Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Understanding Fairness and Favouritism in AI-Based HR Appraisals: A Mixed-Methods Analysis of Employees in the Indian IT Sector

Yajikumar E M
Vellore Institute of Technology University
Sowmya G
sowmya.g@vit.ac.in

Vellore Institute of Technology University

Article

Keywords: Algorithmic Transparency, Perceived Fairness, Workplace Favouritism, Digital Literacy, AI in HR, Algorithmic Management

Posted Date: March 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8847797/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

AI-driven HR systems are increasingly adopted across organisations, yet concerns regarding transparency, fairness, and hidden bias continue to affect employee experiences. This mixed-method study examines how algorithmic transparency influences perceived fairness and workplace favouritism, and how digital literacy moderates these relationships within the Indian IT environments. Qualitative insights from 20 phenomenological interviews further revealed themes of black-box opacity, data visibility bias and algorithmic favouritism, highlighting how contextual sensitivity shapes negative user experiences. Quantitative data from 400 employees were analysed using PLS-SEM, revealing that transparency significantly enhances fairness perceptions, which in turn reduces perceived favouritism, confirming fairness as a key psychological mediator. Digital literacy emerged as a strong moderator, indicating that employees with higher literacy interpret AI outputs more accurately, reducing suspicion and algorithmic anxiety. Triangulating across methods, the study concludes that fairness is not achieved through automation alone. Rather, it arises from transparency, interpretability, and digital empowerment. Findings provide actionable implications for AI governance, HR design, and organisational policy, underscoring the importance of human-centred algorithmic management.

1. Introduction

The rapid growth of artificial intelligence (AI) in human resource management (HRM) has transformed traditional managerial functions by automating decisions that were previously controlled by humans. [1].From candidate screening to performance appraisal. AI systems now mediate organisational judgments about people[2]. However, the black box problem means opacity of algorithmic systems has generated significant concerns about algorithmic fairness, transparency, and perceived favouritism in digital workplaces [3,4].Suspicions of hidden biases and technological favouritism can emerge when employees cannot understand how algorithmic decisions are made [5].In Indian IT sector firms like Infosys, TCS, Wipro etc. are using AI-driven HR systems to determine hiring, promotions, and even wage adjustments [6]. However, transparency in algorithmic decision-making is often limited, leaving employees uncertain about the reasonableness of AI-enabled judgments[7]. Employees' perceptions of favouritism may shift from human to machine bias when organisations deploy their predictive models to identify high performers [8,9]. This requires a deeper understanding of how algorithmic transparency influences employees' perceived fairness of technology and whether this perceived fairness can mitigate workplace favouritism, in the context of digital HR.

Algorithmic transparency refers to the degree to which AI systems disclose their logic, data sources, and decision rules to stakeholders [10]. Transparent systems allow users to understand how inputs lead to outputs, thereby reducing uncertainty and enhancing trust [11]. In HR, transparency is crucial because opaque algorithms can perpetuate hidden biases embedded in training data [12].Research shows that when employees perceive decision-making systems as transparent, they report higher procedural justice and trust in HR outcomes [13,14]. Conversely, a lack of transparency can erode perceived fairness and

foster perceptions of favouritism,[15]. Perceived fairness reflects employees' evaluation of the equity and impartiality of technological systems [16,17]. It encompasses both procedural fairness and distributive fairness. Recent studies suggest that transparency improves fairness by making algorithmic processes auditable and explainable[18]. When employees perceive AI systems as fair, they are less likely to attribute favouritism to organisational processes [19]. Thus, perceived fairness of technology can serve as a mediating mechanism through which algorithmic transparency influences workplace favouritism. While transparency and fairness are important, employees' ability to interpret algorithmic explanations termed digital literacy, modulates these effects [20]. High digital literacy enables users to comprehend algorithmic reasoning and detect inconsistencies[21]. Conversely, low digital literacy may amplify confusion, undermining fairness perceptions even in transparent systems [22]. Therefore, digital literacy is conceptualised as a moderator that strengthens the link between perceived fairness and reduced favouritism. Workplace favouritism traditionally arises when employees perceive that decisions are influenced by personal biases or social ties rather than merit [23].In digital contexts, favouritism manifests through algorithmic bias when certain groups or individuals systematically benefit from model outcomes [24]. For instance, AI tools trained on historical HR data may accidentally favour employees who look like previous high performers [25]. Thus, algorithmic favouritism extends the idea of organisational politics into the algorithmic domain and creates new forms of digital inequality [26]. Although algorithmic fairness and transparency have been examined separately by existing studies,[13,15,18], their combined effect on perceived favouritism in HR contexts has been explored by limited research, particularly in emerging economies like India. Furthermore, few studies integrate both quantitative validation and phenomenological exploration of employees' lived experiences with AI-driven HR systems. This gap underscores the need for a mixed-method investigation that combines empirical modeling with qualitative depth [27,28]. Therefore, this study addresses two research questions.

RQ1

How does algorithmic transparency influence perceived workplace favouritism both directly and indirectly through employees' perceptions of the fairness of technology, and how does this relationship depend on the digital literacy level of employees?

RQ2

How do employees interpret and make sense of their lived experiences with algorithmic human resource systems in shaping perceptions of fairness and favouritism in the workplace?

Employing an exploratory sequential mixed-methods approach, this research combines the qualitative findings from interpretive interviews with quantitative hypothesis testing using Structural Equation Modeling (SEM). This research can enhance the literature on algorithmic decision-making within organizations by revealing how fairness perceptions of AI-driven decision-making systems serve as a fundamental psychological process. It reinterprets work-related favouritism as a technology-driven

outcome and highlights the influence of digital literacy and trust in technology as moderating factors in the relationship between algorithmic transparency, fairness perceptions, and favouritism.

This paper is organised as follows. It first reviews literature on algorithmic transparency, AI-based decision-making, perceived fairness, and workplace favouritism. The qualitative phase is then presented to capture employees' experiences with algorithmic systems and identify key themes. From these insights, hypotheses are formulated and examined through quantitative methods. This quantitative phase encompasses measurement and structural model analysis utilising SEM, as well as moderation analysis focusing on digital literacy and trust in technology. The paper concludes with discussions on implications, limitations, and suggestions for future research.

2. Literature Review

Previous research consistently emphasises that transparency is a crucial driver of user fairness and trust in algorithmic decision-making environments [18,29]. Prior researchers highlights that algorithmic opacity, data visibility bias, and contextual insensitivity may generate new forms of workplace favouritism[30,31]. Additionally, research indicates that digital literacy has a significant impact on how workers understand algorithmic assessments, lowering bias attributions and moderating perceived fairness. [32]. These studies highlight a clear research gap on the relationship between transparency, equity, and digital literacy in HR systems based on AI, particularly within the Indian IT sector. Table 1 provides a comprehensive summary of key studies, detailing their aims, findings, and relevance to the current research model. These studies provide the theoretical foundation for the hypothesised relationships and highlights the need for a mixed-method study on employee experiences with AI-driven HR systems.

Table 1
Comprehensive summary of key studies

Author	Objectives	Key Findings	Relevance to Current Study
Colquitt [33]	The study aimed to clarify and validate the multidimensional nature of organisational justice. It focuses on how employees form fairness perceptions based on procedures and outcomes.	The study found that trust, acceptability, and compliance are all highly predicted by procedural justice. It also made clear that fairness assessments are more influenced by the consistency and transparency of processes than by results alone.	This work provides a theoretical foundation for linking algorithmic transparency to fairness. It supports the idea that clear and consistent AI logic increases fairness perceptions in algorithmic HR systems.
Bies [34]	This paper examines how perceptions of fairness shape reactions to organisational decision-making. It specifically focuses on how explanations, respect, and information sharing influence bias attributions.	Findings show that fairness perceptions significantly reduce feelings of bias and favouritism, The study also argues that explanatory depth influences whether employees view decisions as legitimate.	This directly supports the fairness → reduced favouritism, pathway in the study. It also highlights the importance of explanations in reducing perceived bias, mirroring in qualitative themes.
Burrell [4]	The paper explores why algorithmic systems are opaque and how this opacity affects user trust and accountability. It categorises opacity into intentional, illiterate, and intrinsic types	The study argues that opacity prevents users from understanding, questioning, and trusting algorithmic decisions. Also shows that black-box systems produce anxiety and suspicion among users.	This strongly explains qualitative findings on “black-box anxiety.” It supports the central role of algorithmic transparency in reducing confusion and mistrust.
Selbst & Barocas [35]	Main objective of the paper is to challenge the idea that AI is fair when it is explained as little as possible and this paper emphasize that fairness requires contextual and ethical reasoning, not just mechanical transparency.	They find that narrow explanations may mislead users and damage trust. The study argues that fairness needs richer human-centric explanations integrated with contextual detail.	This supports the finding that employees want contextual fairness, not just numeric outputs. It reinforces the need for meaningful explanations in AI-driven HR evaluations.

Author	Objectives	Key Findings	Relevance to Current Study
O'Neil [36]	This book examines how algorithms can reproduce inequality by embedding historical and structural bias. It explores how data visibility shapes who gets rewarded or penalized.	The author finds that systems often favor roles with high digital footprints while ignoring invisible labor. These forms of bias occur even when algorithms appear neutral.	This directly supports the emergent theme of algorithmic favouritism, / visibility bias in this qualitative findings. It validates the conclusion that favouritism, can be "data-driven," not human-driven.
Lee [37]	The study examines how users interpret algorithmic fairness based on explainability and information accessibility. It also explores how explanations influence trust.	Findings indicate that transparent algorithms increase perceived legitimacy and acceptance. The study also shows that explanations reduce perceptions of bias and promote fairness.	Supports AT → PF relationship. It validates the need for explanation dashboards suggested in practical implications.
Mittelstadt [38]	The paper analyses ethical risks in algorithmic decision-making and suggests fairness standards that go beyond accuracy.	It shows that fairness is dependent on context, human expectations, and interpretability. The study highlights that equal rules may still produce unfair results.	Aligns with qualitative theme of contextual fairness. Supports argument that AI treating everyone equally can still feel unfair when context is ignored.
Wang & Siau [39]	This review examines factors that build trust in AI systems, including transparency, performance, and user understanding. It integrates AI trust research with ethical considerations.	The study finds that transparency improves trust by reducing uncertainty and also shows that users' ability to interpret AI outputs shapes their perception of fairness.	Supports pathways: Fairness → Trust and DL moderates fairness. It reinforces the importance of user comprehension.
Long & Magerko[40]	The paper provides a conceptual framework for AI literacy, defining key competencies required for understanding algorithmic systems. It investigates how literacy affects user experiences.	The authors show that literacy decreases fear and anxiety while increasing proper trust. They also argue that literacy improves users' interpretation of algorithmic outputs.	Directly supports the moderating effect of digital literacy found in our model. In qualitative phase, it explains why high DL employees reported less favouritism.
Dogruel et al. [41]	This study aims to conceptualize algorithmic literacy and develop its measurement dimensions. It focuses on how understanding AI systems influences user attitudes and critical reflections.	Findings show that algorithmic literacy significantly reduces fear, suspicion, and misinformation regarding AI systems. It also enhances users' ability to interpret results and evaluate fairness.	Strongly supports the moderating role of digital literacy in the study. Helps explain why high-DL employees interpret AI evaluations more positively and perceive less favouritism,

Author	Objectives	Key Findings	Relevance to Current Study
Meijerink et al.[2]	This study looks at how employee acceptance and perceptions of fairness are influenced by HR analytics systems. It highlights the psychological mechanisms through which analytics-driven decisions influence trust.	It is revealed by the study that clarity, transparency, and communication are essential for employees to accept data-driven evaluations.	Directly supports AT → PF → WF pathway. Provides HR-specific evidence that transparency and fairness determine acceptance of algorithmic decisions.
Shin[42]	This study investigates how different levels of AI transparency influence trust, fairness perception, and user acceptance. It incorporates behavioral experiments with AI systems.	Results show that higher transparency significantly improves interpretability, fairness, and trust. Transparent AI reduces perceived bias and enhances legitimacy.	Fully supports central hypothesis: Transparency enhances fairness, which reduces favouritism,. Reinforces the idea that explainability builds trust.
Nicolas et al.[43]	The aim is to understand how explainable AI affects worker trust and acceptance in algorithmic management environments.	The study reverled that explanation mechanisms reduce bias concerns, increase feelings of fairness, and improve user empowerment.	Supports the findings that employees want participatory fairness and training. Reinforces the combined role of Transparency + Digital Literacy.
Rohden et al.[44]	This paper explores how workers emotionally respond to algorithmic management. It analyses strategies employees use to cope with algorithmic evaluation.	Findings indicate that opacity leads to anxiety, fear, and strategic adaptation. Workers experience both increased scrutiny and reduced autonomy.	Confirms qualitative themes such as anxiety, adaptation, and conditional trust in study.
Eiband et al.[45]	The aim is to study how different types of explanation interfaces influence transparency and user trust.	Findings show that incomplete or vague explanations can backfire and reduce trust. Clear, actionable explanations increase acceptance.	Supports recommendation to develop transparency dashboards with interpretive clarity.
Kizilcec[46]	This study examines the optimal level of algorithmic explanation required to build trust without overwhelming users.	It finds that users prefer concise but meaningful explanations and that unnecessary complexity harms trust.	Supports claim that transparency must be contextual and digestible, not overly technical.
Venkatesh et al.[47]	The study expands IT adoption theory by highlighting the role of user skills, facilitating	It finds that competence and confidence significantly predict technology acceptance and satisfaction.	Provides strong theoretical justification for including Digital Literacy as a moderator in the model.

Author	Objectives	Key Findings	Relevance to Current Study
	conditions, and perceived usefulness.		
Mayer[48]	The study examines how users cognitively process explanations in complex systems.	It demonstrates that explanations reduce cognitive load and increase comprehension.	Supports the proposition that explanations reduce anxiety and improve fairness perceptions.
Alaimo et al.[49]	The study explores how digital systems make certain forms of work more visible than others.	It shows that digital traces shape what organizations see as valuable, influencing performance assessment.	Strongly reinforces the theme of data visibility bias and algorithmic favouritism,

3 Theoretical Framework

This study is theoretically informed by Organisational Justice Theory (OJT), which offers a robust framework for examining how employees interpret fairness, bias, and favouritism in technology-mediated organisational contexts. Organisational justice theory was used in this study as a sensitising theory, which means that it was used to inform variable selection and interpretation but did not predetermine findings. According to organisational justice theory, employee attitudes and behaviours are shaped less by objective organisational practices and more by subjective evaluations of fairness related to decision-making processes, information disclosure, and outcome distributions[33,50]. As algorithmic and AI-driven systems increasingly mediate organisational decisions, these justice evaluations become central to how employees make sense of technology-enabled management practices.

Algorithmic transparency is conceptualised as the independent construct representing the extent to which employees can access, understand, and interpret the logic in organisational decisions. Organisational justice theory explains that transparent procedures allow individuals to assess the neutrality, consistency, and ethicality of decision rules, which are foundational factors of procedural and informational justice [51]. Prior research suggests that opaque or black-box decision systems can create uncertainty, suspicion, and perceived injustice, even when outcomes appear objectively neutral [30]. Accordingly, this study treats algorithmic transparency as a critical organisational condition that may shape how employees evaluate fairness in AI-mediated decision environments. According to the exploratory nature of the study, it is not believed that transparency would always result in a more favorable view of fairness.Rather, its function is investigated experimentally to learn how workers perceive and react to algorithmic exposure.

The perceived fairness of technology has been established as a central explanatory construct reflecting employees' fair evaluations of AI-controlled systems. Derived from organisational justice theory, fairness perceptions are understood in a multifaceted way. And it is a part of procedural fairness, informational fairness, and distributive fairness[33,34].Rather than assuming that technology is inherently fair or biased,

this study conceptualises perceived fairness as a cognitive and interpretive judgment formed through employees' interactions with algorithmic systems. Perceived fairness is a crucial lens through which algorithmic activities are regarded since existing research indicates that workers frequently use fairness heuristics when assessing complicated or poorly understood systems[52]. In this exploratory framework, perceived fairness of technology is treated as a potential mechanism through which algorithmic practices may influence broader organisational perceptions, without assuming a universal or stable pathway

The dependent construct, perceived workplace favouritism, emphasises employees' beliefs that organisational decisions and rewards are influenced by personal relationships, biased treatment, and managerial bias rather than merit-based criteria. Within organisational justice theory, favouritism is commonly interpreted as a symbolic indicator of justice violations[53]. Importantly, favouritism in this study is conceptualised as a perceptual and socially constructed phenomenon. Justice research regularly shows that unfavourable workplace attitudes and sensemaking processes are driven by perceptions of injustice rather than real disparity alone[54]. From an exploratory point of view, this study investigates how employees come to associate algorithmic decision systems with favouritism perceptions, especially in the contexts of technology-driven work environments.

Digital literacy is incorporated as a moderating condition that may shape how employees interpret algorithmic transparency and fairness cues. Organisational justice theory recognises that fairness evaluations are contingent upon individuals' capacity to understand and evaluate organisational procedures[55]. In technology-mediated contexts, digital literacy reflects employees' ability to comprehend algorithmic logic, data-driven reasoning, and system limitations. Employees with higher levels of digital literacy may be better to interpret algorithmic explanations and differentiate procedural rigour and managerial manipulation. On the opposite side, lower levels of digital literacy may intensify ambiguity, reduce perceived controllability, and increase sensitivity to potential injustice, even when transparency mechanisms are assured [56]. In line with the exploratory study, digital literacy is treated as a boundary condition whose role is empirically examined rather than theoretically assumed.

Taken together, the proposed framework conceptualizes algorithmic decision-making as a justice-relevant organisational practice whose implications for workplace favouritism are shaped by employees' fairness interpretations and interpretive capacities. By grounding the study in organisational justice theory while maintaining an exploratory stance, the framework allows for the identification of emergent patterns, alternative mechanisms, and contextual nuances through sequential qualitative and quantitative inquiry. This approach enables theory refinement rather than theory confirmation, aligning closely with the study's mixed-methods design and exploratory objectives.

4. Exploratory study: Phase 1

4.1 Interpretive Phenomenological Study

4.1.1 Methodology

The qualitative design using Interpretive Phenomenological Analysis (IPA) has used to explore employees' lived experiences of algorithmic transparency, fairness, and favouritism in AI-enabled HR systems. A snowball sampling technique was used to recruit 20 employees from Indian IT firms who have direct exposure to AI-based HR applications. Semi-structured interviews were conducted over 15 days, with each interview lasting approximately 30–45 minutes. Written informed consent was obtained before participation and audio recording. In addition to recordings, field notes were maintained to capture emotional tone, pauses, and contextual cues. Interview recordings were transcribed verbatim and analysed using QualCoder, following established IPA analytic procedures. Analysis began with repeated reading of transcripts to achieve immersion and idiographic focus. Initial exploratory coding captured descriptive, linguistic, and conceptual aspects of participants' accounts. These were subsequently transformed into emergent codes and organised into a structured codebook comprising code number, code name, operational definition, illustrative quotes, and corresponding themes. Through iterative comparison across cases, related codes were clustered into higher-order themes that reflected shared experiential patterns while preserving sensitivity to individual meaning-making. To ensure analytic rigour, emergent themes were continuously cross-checked against original transcripts and field notes. Coding credibility was supported through independent review by co-authors and two external qualitative researchers, with inter-coder agreement exceeding recommended qualitative consistency thresholds [57]. Any discrepancies were resolved through discussion and reference to the original data.

This study involved voluntary participation of adult employees from the Indian IT sector and did not involve clinical procedures, medical interventions, or vulnerable populations. In accordance with institutional research guidelines, anonymous and non-interventional survey-based social science research does not require formal ethics committee approval.Prior to participation, respondents were informed about the purpose of the study, the voluntary nature of participation, and their right to withdraw at any time without consequence. Informed consent was obtained electronically before data collection commenced. No personally identifiable information was collected or stored, and all responses were anonymized prior to analysis. All methods were performed in accordance with relevant institutional guidelines and regulations and adhered to internationally accepted ethical principles for research involving human participants, including the Declaration of Helsinki

4.1.2 Qualitative Analysis

The qualitative phase of this study, based on 20 in-depth phenomenological interviews, explored employees' lived experiences with AI-driven HR systems in Indian IT organisations. Seven main themes emerged from the analysis: Algorithmic Transparency, Algorithmic Fairness, Algorithmic Favouritism, Digital Literacy, Trust in AI Systems, Emotional and Behavioural Responses, and Expectations for Transparency Improvement, which collectively illustrate how employees interpret and navigate AI-enabled HR decision-making. The prevalence and depth of these themes demonstrate strong data saturation, as evidenced by the frequency distribution presented in Table 2.

Table 2: Code Tree Frequency Table

Code Tree	Total
1 Algorithmic Transparency	**45**
1.1 Clarity of Decision Logic	
Code 1	17
1.2 Visibility of Data Inputs	
Code 2	13
1.3 Explanation of Outcomes	
Code 3	15
2 Algorithmic Fairness	**28**
2.1 Equal vs Contextual Fairness	
Code 4	13
2.2 Accountability of System	
Code 5	8
2.3 Alignment with Human Ethics	
Code 6	7
3 Algorithmic Favouritism	29
3.1 Role or Data Bias	
Code 7	11
3.2 Perceived Unfair Advantage	
Code 8	7
3.3 Contextual Blind Spots	
Code 9	11
4 Digital Literacy	**39**
4.1 Ability to Interpret Results |	
Code 10	17
4.2 Confidence in Using AI Tools	
Code 11	12
4.3 Awareness of Algorithmic Limits	
Code 12	10
5 Trust in AI-Driven Systems	**33**
5.1 Initial Skepticism	
Code 13	10
5.2 Conditional/Pragmatic Trust	
Code 14	18
5.3 Trust Through Familiarity	
Code 15	5
6 Emotional & Behavioural Responses	**38**
6.1 Anxiety/Uncertainty	
Code 16	17
6.2 Adaptation & Acceptance	
Code 17	11
6.3 Resistance & Disengagement	
Code 18	10
7 Expectations for Transparency & Improvement	**40**
7.1 Desire for Feedback/Communication	
Code 19	13
7.2 Participatory Fairness	
Code 20	17

7.3 Need for Training/Involvement	
Code 21	10

One of the most prevalent topics was algorithmic transparency, with employee regularly bringing up the opaqueness of AI decision-making, with subthemes such as Clarity of Decision Logic, Visibility of Data Inputs, and Explanation of Outcomes appeared repeatedly. Participants consistently expressed uncertainty regarding how performance scores were generated, often describing the system as a black box. This pattern aligns with the core quantitative finding that algorithmic transparency significantly shapes perceptions of fairness. The high frequency of these codes, visually reinforced by the expansive segments in Fig. 2, highlights that transparency concerns are foundational to employee experiences.

Algorithmic Fairness was similarly influential, reflected in 28 coding instances. Respondents perceived the AI system as equal in procedure but unequal in effect, particularly due to its inability to account for contextual factors such as task complexity or role-specific demands. This “mechanical fairness” was contrasted with human ethical judgment, reinforcing quantitative evidence that fairness mediates the relationship between transparency and workplace favouritism. Subthemes like Equal vs. Contextual Fairness and Alignment with Human Ethics were frequently noted, indicating that employees assess fairness through both outcome and ethical lenses. Algorithmic Favouritism appeared in 29 coding instances and centred on perceptions of data-driven bias. Participants described how roles with high digital visibility received higher AI recognition compared to roles involving intangible contributions. Subthemes such as Role or Data Bias and Contextual Blind Spots illustrate how favouritism emerges from what the algorithm can or cannot see. Digital Literacy was one of the most critical themes with 39 coding instances. Participants reported feeling more fair and less worried about favouritism when they got a better understanding of how the AI system collected and analysed data. Subthemes such as Ability to Interpret Results and Confidence in Using AI Tools were strongly expressed. The prominence of this theme in Fig. 1 clearly shows its central role in shaping employee experiences. Trust in AI-Driven Systems emerged as a conditional and dynamic theme with 33 coding instances. Employees initially approached AI with doubt but developed conditional trust through continuous exposure and understanding. Trust increased not because the AI was perfect, but because employees learned its logic of implementation. The quantitative path showing transparency influences fairness, which in turn fosters lower favouritism and higher trust. Another main trend, Emotional and Behavioural Responses, also featured prominently, with 38 coding instances. Employees described emotions ranging from confusion and anxiety to adaptation and acceptance. The Anxiety/Uncertainty subtheme appeared frequently, emphasising how opacity leads to emotional strain. As participants gained confidence through digital literacy, anxiety reduced, and adaptive behaviours emerged such as improved documentation, increased visibility, and intentional alignment with algorithmic expectations. Finally, Expectations for Transparency and System Improvement highlighted employees’ desire for participatory fairness. Participants emphasised the need for interpretive dashboards, clear communication, feedback mechanisms, and training. They argued that AI cannot be fair without human collaboration.

Table 3
Codes, Example quote and Thames

Code No.	Code Name (Sub Theme)	Definition	Example Quote	Themes
Code 1	Clarity of Decision Logic	Employees' understanding of how AI calculates or assigns performance scores.	"The system gives a number, but I still don't know how it decides what matters more."	Algorithmic Transparency
Code 2	Visibility of Data Inputs	Awareness of what data the AI system captures and values.	"It tracks logins and closures, but not the creative part of my work."	
Code 3	Explanation of Outcomes	Extent to which the AI explains final ratings or evaluations.	"I got my score, but there was no explanation of what led to it."	
Code 4	Equal vs Contextual Fairness	Perception of whether AI treats everyone equally while ignoring contextual factors.	"It uses the same formula for all roles, even though our work is completely different."	Experience of Algorithmic Fairness
Code 5	Accountability of System	Whether employees believe the organization takes responsibility for AI-driven decisions.	"When the AI gets it wrong, no one takes accountability it's just blamed on the system."	
Code 6	Alignment With Human Ethics	Comparison between algorithmic fairness and human ethical judgment.	"The system is fair mathematically, but not morally."	
Code 7	Role or Data Bias	Structural favouritism caused by visibility differences in job roles.	"People whose work leaves more digital traces always get better scores."	Algorithmic Favouritism
Code 8	Perceived Unfair Advantage	Belief that certain behaviors or roles are rewarded unfairly by the AI.	"Just updating tickets frequently boosts your rating, even if the work itself isn't impactful."	
Code 9	Contextual Blind Spots	AI's inability to capture qualitative contributions or	"It doesn't pick up mentoring or the complexity of the	

Code No.	Code Name (Sub Theme)	Definition	Example Quote	Themes
		exceptional circumstances.	backend issues I solve."	
Code 10	Ability to Interpret Results	Skill in understanding AI dashboards and score patterns.	"Once I learned how the metrics work, my anxiety reduced."	Digital Literacy
Code 11	Confidence in Using AI Tools	Self-efficacy developed through interacting with AI systems.	"After a few cycles, I became more confident navigating the system."	
Code 12	Awareness of Algorithmic Limits	Recognition that AI systems are not perfect or context-sensitive.	"The AI is consistent, but it cannot understand the full complexity of my job."	
Code 13	Initial Skepticism	Early doubts or mistrust toward AI-based evaluations.	"At first, I didn't believe a machine could evaluate us fairly."	Trust in AI Driven Systems
Code 14	Conditional/Pragmatic Trust	Trust built on predictability, not on belief that the system is fully fair.	"I trust it to be consistent, but not to be complete."	
Code 15	Trust Through Familiarity	Growing trust formed through repeated use and understanding.	"The more I used it, the more predictable and less threatening it felt."	
Code 16	Anxiety/Uncertainty	Stress or confusion caused by opaque AI decisions.	"I used to get anxious every time a new score appeared."	Emotional & Behavioral Responses
Code 17	Adaptation & Acceptance	Behavioral adjustments employees make to align with AI expectations.	"I started documenting everything so the system could finally 'see' my work."	
Code 18	Resistance & Disengagement	Withdrawal or demotivation caused by perceived unfairness.	"At one point, I stopped caring because the system never recognized my real contributions."	

Code No.	Code Name (Sub Theme)	Definition	Example Quote	Themes
Code 19	Desire for Feedback/Communication	Need for transparent explanations and two-way communication.	“If they just explained how the score was generated, things would feel fairer.”	Expectations for Transparency & Improvement
Code 20	Participatory Fairness	Desire to be included in decisions about AI design or metric settings.	“We should be involved in defining what counts as performance, not just judged by it.”	
Code 21	Need for Training/Involvement	Expectation for training to build AI literacy and reduce confusion.	“With proper training, we’d understand the system instead of fearing it.”	

Insights from Table 3, Fig. 1, and Fig. 2 taken together show substantial agreement between qualitative phases and consistent thematic saturation.The qualitative narratives add depth to the quantitative model, showing that employees do not merely react to AI outcomes they interpret, negotiate, and emotionally respond to them through the lenses of transparency, fairness, literacy, and trust.

5. Hypothesis Development

This study follows an exploratory sequential mixed-methods design, where hypotheses are developed after completing the qualitative study. Using interpretative phenomenological analysis, the qualitative phase helped identify important trends in how Indian IT workers perceive AI-driven decision systems, particularly about algorithmic transparency, perceived technological fairness, and workplace favouritism.[27].

Based on these qualitative trends and the theoretical framework discussed earlier, the following hypotheses were formulated for quantitative testing. This approach ensures that the hypotheses are grounded in real experiences of employees and follow accepted mixed-methods research practices [58].

5.1 Algorithmic Transparency and Perceived Workplace Favouritism.

Algorithmic transparency refers that, which employees can understand how AI-driven decision systems generate workplace outcomes. As algorithmic systems increasingly mediate managerial decisions may heighten employee uncertainty regarding hidden discretion or preferential treatment[2] .According to

earlier studies, transparent systems convey consistency and rule-based decision-making, which allays fears of favoritism, whereas opaque algorithmic methods enhance perceived bias and skepticism[59].

The increasing use of AI-based decision systems in organisations has changed how employees interpret managerial authority and accountability. In technology-mediated environments, employees often lack direct visibility into how decisions are produced. When algorithmic processes are unclear, employees may attribute outcomes to hidden human intervention, biased data inputs, or selective manipulation, particularly in contexts where career outcomes such as promotions, performance evaluation and appraisals are involved. More than that previous research suggests that reduced transparency increases ambiguity, which in turn encourages employees to rely on social interpretations, including assumptions of favouritism or preferential treatment [60,61]. Therefore, by making it clear how results are produced, transparency acts as a systemic mechanism that restricts speculative attributions.

H1

Algorithmic transparency in AI-driven decision systems is negatively associated with perceived workplace favouritism.

5.2 Algorithmic Transparency and Perceived fairness of technology

Perceived fairness of technology reflects employees' evaluations of whether AI systems operate impartially, consistently, and without bias. Transparency enables employees to assess decision logic and evaluation criteria, strengthening procedural fairness perceptions [33]. Empirical evidence indicates that explainable and transparent AI systems enhance perceived fairness, trust, and acceptance in organisational contexts [42].

Employees' evaluations of AI-driven systems are shaped not only by outcomes but also by their understanding of how those outcomes are produced. Transparency allows employees to examine decision rules, data sources, and evaluative criteria, thereby facilitating judgments about whether systems operate consistently and impartially. In the absence of such clarity, employees may perceive algorithmic decisions as arbitrary or biased, even when they are technically accurate. Prior studies on explainable AI indicate that transparency enhances perceptions of procedural justice by enabling individuals to assess whether similar inputs yield similar outcomes[42]. As a result, transparency becomes a foundational condition for fairness evaluations in algorithmic decision environments.

Perceived fairness of technology reflects employees' evaluations of whether AI systems operate impartially, consistently, and without bias. Transparency enables employees to assess decision logic and evaluation criteria, strengthening procedural fairness perceptions [33]. Empirical evidence indicates that explainable and transparent AI systems enhance perceived fairness, trust, and acceptance in organisational contexts [42,62].

H2

Algorithmic transparency in AI-driven decision systems is positively associated with perceived fairness of technology.

5.3 Perceived fairness of technology and Perceived workplace favouritism

Perceived workplace favouritism arises when employees believe that organisational decisions are influenced by subjective or relational factors rather than merit-based criteria [23]. When AI systems are perceived as unfair or biased, employees may infer that favouritism persists in a technologically mediated form. Research consistently shows that perceptions of unfair procedures increase favouritism beliefs, distrust, and organisational cynicism [63].

Perceptions of favouritism often emerge when employees believe that organisational systems fail to apply rules uniformly. In algorithmic contexts, fairness perceptions become particularly salient because employees may lack opportunities to contest or question automated decisions. When AI systems are perceived as unfair, employees may conclude that favouritism persists in indirect or concealed forms, such as biased data training, selective overrides, or unequal system calibration. Research suggests that unfair procedures whether enacted by humans or technologies undermine legitimacy and encourage employees to attribute outcomes to favouritism rather than merit [63]. Thus, fairness perceptions play a central role in shaping whether employees view AI as neutral or as an extension of biased organisational practices.

Perceived workplace favouritism arises when employees believe that organisational decisions are influenced by subjective or relational factors rather than merit-based criteria [64]. When AI systems are perceived as unfair or biased, employees may infer that favouritism persists in a technologically mediated firm. Research consistently shows that perceptions of unfair procedures increase favouritism beliefs, distrust, and organisational cynicism [63].

H3

Perceived fairness of technology is negatively associated with perceived workplace favouritism.

5.4 Mediating Role of Perceived Fairness of Technology

Perceived fairness of technology is expected to function as a key explanatory mechanism linking algorithmic transparency and favouritism perceptions. Studies on algorithmic governance suggest that fairness perceptions represent the primary pathway through which transparent and explainable AI systems gain legitimacy and acceptance [65].

Perceived fairness of technology is expected to function as a key explanatory mechanism linking algorithmic transparency and favouritism perceptions. Transparency alone may not reduce favouritism unless it meaningfully enhances fairness judgments. Prior research in algorithmic governance highlights that fairness perceptions serve as a psychological mechanism through which transparency influences acceptance and legitimacy. Without improved fairness judgments, transparency may remain symbolic rather than substantive. Therefore, perceived fairness of technology is expected to function as the critical interpretive link that transforms transparency into reduced favouritism attributions[65].

H4

Perceived fairness of technology mediates the relationship between algorithmic transparency and perceived workplace favouritism.

5.5 Moderating Role of Digital Literacy

Digital literacy reflects employees' ability to interpret and critically evaluate AI-based systems. Employees vary substantially in their ability to interpret and critically evaluate algorithmic information. Employees with higher digital literacy are more capable of interpreting algorithmic logic, questioning data inputs, and distinguishing between system bias and legitimate rule-based decisions. In contrast, employees with lower digital literacy may experience cognitive overload, limiting the effectiveness of transparency initiatives[66]. Empirical studies suggest that digital competence moderates employees' responses to AI systems, influencing perceptions of fairness and resistance. And another prior research demonstrates that digital competence conditions employees' responses to AI systems, influencing fairness perceptions, resistance, and acceptance [67].

H5

Digital literacy moderates the relationship between algorithmic transparency and perceived fairness of technology, such that the relationship is stronger at higher levels of digital literacy.

Therefore, in this context, the theoretical model has been formulated and is illustrated in Fig. 3

6. Methodology phase 2 (Quantitative study)

6.1 Methods

The quantitative phase was conducted after the qualitative study, following an exploratory sequential mixed-methods design, to test the hypotheses developed from the qualitative findings [27]. The study targeted employees working in IT organisations in India, where AI-driven decision systems are commonly used for workplace decisions. Data were collected using a structured questionnaire administered online between May 2025 and August 2025. India was selected as the research context due to the increasing use of algorithmic systems in IT workplaces. To ensure relevance, two screening questions were used:

(1) Does your organisation use digital or AI-based systems for workplace decisions such as performance evaluation, task allocation, monitoring, or promotion? (2) Have you personally been affected by at least one decision made or supported by such systems in your current organisation? Only respondents who answered “yes” to both questions were included. A purposive sampling method was adopted to ensure that respondents had the required familiarity with AI-based HR systems. Data were collected from professionals working at major IT firms in India through online structured questionnaires distributed via Google Forms. Out of 450 questionnaires distributed, 400 valid responses were retained for the final analysis after screening for completeness and consistency.

6.2 The measurement instrument

A self-administered, structured questionnaire with questions from validated measures to assure construct and content validity served as the study tool. All items were measured using a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The questionnaire comprised four constructs and was adapted from the previously established scales given below.

Algorithmic Transparency (AT)

Adapted from Shin[42] and Lepri et al. [68], focusing on the clarity and explainability of AI-driven HR processes.

Perceived Fairness of Technology (PF)

Drawn from Colquitt et al[17] and Shin[42] measuring procedural and distributive fairness perceptions.

Workplace Favouritism (WF)

Adapted from Khatri and Tsang[23] and Meijerink et al[2] assessing perceived partiality or algorithmic bias in HR decisions.

Digital Literacy (DL)

Based on Ng[20] and Van Deursen and Van Dijk [22], measuring an individual’s competence in understanding and critically evaluating digital technologies.

All items underwent expert validation by 2 HR experts to ensure content relevance and face validity [69]. Minor linguistic refinements were made to contextualise items for Indian IT professionals. Before the main data collection, a pilot study with 80 respondents was conducted to test the clarity, reliability, and validity of the instrument. The pilot data allowed refinement of the measurement model by identifying and eliminating low-performing items [70].

The pilot phase revealed strong reliability but indicated certain items with low outer loadings and cross-loading concerns. Following the guidelines of Hair et al.[70], items with loadings below 0.70 were dropped. Constructs initially contained 24 items, but after refinement, 16 items were retained across four

constructs. Some items were dropped (AT3, AT5, AT7, PF1, PF4, PF8, WF3, WF5, WF8, DL2, DL6, DL7) because of showing redundancy or conceptual overlap. Their removal helped to improve discriminant validity, with all Average Variance Extracted (AVE) values above 0.50 and Cronbach's alpha values above 0.80. The revised instrument achieved Composite Reliability (CR) ranging from 0.89 to 0.93, demonstrating internal consistency [71].

Table 4
Factor loadings.

Construct	Code	Item Statement	Factor Loading
Algorithmic Transparency (AT)	AT1	The AI-based HR system clearly explains how employee evaluations are conducted.	0.88
	AT2	I can understand how the system uses my data to make HR decisions.	0.85
	AT4	The decision-making process of the AI system is open and visible.	0.82
	AT6	The rules behind AI decisions are communicated transparently.	0.81
Perceived Fairness of Technology (PF)	PF2	The AI system treats all employees fairly.	0.83
	PF3	The AI-based HR process is unbiased and consistent.	0.8
	PF5	The organization uses AI to promote fairness in HR.	0.82
	PF6	AI reduces favouritism compared to human judgments.	0.78
	PF7	The AI system's data handling appears fair and ethical.	0.8
Workplace Favouritism (WF)	WF1	Some employees benefit more from AI-based HR systems than others.	0.83
	WF2	The AI system favors certain employee types.	0.82
	WF4	AI tools create hidden advantages for specific profiles.	0.8
	WF6	AI results sometimes reflect human-like biases.	0.78
	WF7	Even with automation, favouritism still exists.	0.79
Digital Literacy (DL)	DL1	I can interpret reports produced by AI-based HR systems.	0.84
	DL3	I understand how AI-based HR tools work conceptually.	0.81
	DL4	I can evaluate whether AI outcomes are fair or biased.	0.8
	DL5	I know how to check for potential AI errors in HR decisions.	0.78

6.3 Psychometric properties of the measurement model

Internal consistency reliability was verified using Cronbach's alpha, Composite Reliability (CR), and rho_a, all above 0.96, far above the acceptable 0.70 threshold [72]. As summarized in Table 5, α ranged from 0.963 to 0.967, and CR (rho_c) ranged from 0.973 to 0.976, confirming strong internal consistency without redundancy. The Average Variance Extracted (AVE) values also above the 0.50 benchmark [71], contributing to establishing excellent convergent validity.

Table 5
Internal Consistency Reliability

	Cronbach's alpha	Composite reliability (rho_a)	Composite reliability (rho_c)	Average variance extracted (AVE)
AT	0.967	0.967	0.976	0.909
DL	0.963	0.981	0.973	0.9
PF	0.967	0.967	0.974	0.884
WF	0.965	0.966	0.973	0.878

Through the Fornell–Larcker criterion and the HTMT ratio, discriminant validity was acquired. The square root of each construct's AVE was higher than its correlations with other constructs, confirming discriminant validity. HTMT ratios were well below 0.85, suggesting that all constructs are conceptually distinct yet theoretically related.

Table 6
HTMT Analysis

	AT	DL	PF	WF
AT				
DL	0.121			
PF	0.678	0.068		
WF	0.436	0.134	0.533	
DL x PF	0.066	0.016	0.013	0.221

Table 7
Fornell–Larcker Criterion

Constructs	AT	DL	PF	WF
AT	0.954			
DL	−0.116	0.949		
PF	0.655	−0.063	0.94	
WF	−0.422	−0.133	−0.516	0.94

The Variance Inflation Factor (VIF) values for all constructs were below 3.0AT (1.45), PF (1.38), DL (1.12) confirming the absence of collinearity issues (Hair et al., 2021). Low VIF values ensure that predictor constructs in the structural model are independent and yield stable path coefficients. All reliability and validity metrics surpassed the recommended standards: α > 0.70, CR > 0.70, AVE > 0.50, HTMT < 0.85, and VIF < 3. These results confirm that the measurement model is statistically robust and theoretically coherent. Consequently, the constructs are internally consistent, valid, and suitable for hypothesis testing in the structural model.

6.4 Results for Structural Model

The structural model was tested using PLS-SEM with bootstrapping to examine the relationships among AT, PF, WF, DL. As shown in Table X and Figure X, algorithmic transparency had a strong positive effect on perceived fairness of technology (β = 0.655, *t* = 22.238, *p* < 0.001), supporting H2. Algorithmic transparency also showed a significant negative effect on workplace favouritism (β = −0.144, *t* = 2.593, *p* < 0.01), supporting H1.

Perceived fairness of technology strongly reduced workplace favouritism (β = −0.434, *t* = 7.908, *p* < 0.001), confirming H3. Digital literacy also had a significant negative direct effect on workplace favouritism (β = −0.180, *t* = 4.329, *p* < 0.001). Further, the interaction between digital literacy and perceived fairness was significant (β = −0.208, *t* = 5.354, *p* < 0.001), indicating that digital literacy strengthens the effect of fairness in reducing workplace favouritism, thus supporting H5.

Overall, the findings support the proposed structural model and highlight the key role of perceived fairness of technology in explaining how algorithmic transparency and digital literacy reduce perceptions of workplace favouritism.

6.5 Mediation Analysis

The mediation analysis examined the indirect effect of AT on WF through PF. The path AT → PF (β = 0.655, t = 22.238, p < 0.001) was strongly positive, while PF → WF (β = − 0.434, t = 7.908, p < 0.001) was significantly negative. The indirect effect (AT → PF → WF) was thus significant, confirming that higher algorithmic transparency enhances perceived fairness, which subsequently reduces perceptions of favouritism in the workplace.The direct path AT → WF (β = − 0.144, p = 0.01) remained significant but

weaker in context, indicating partial mediation rather than full mediation. This suggests that while transparency directly reduces favouritism perceptions to some extent.

Table 8
Hypothesis Testing

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics	P value	Decisions
AT -> PF	0.655	0.656	0.029	22.238	0.000	**Strong positive;** transparency significantly increases perceived fairness.
AT->WF	-0.144	-0.143	0.056	2.593	0.001	**Strong negative;** higher fairness reduces perceived favouritism.
DL > WF	-0.18	-0.183	0.042	4.329	0.000	**Significant negative effect;** higher digital literacy reduces favouritism.
DL x PF -> WF	-0.208	-0.207	0.039	5.354	0.000	**Significant moderation;** digital literacy strengthens fairness – favouritism relationship.
PF > WF	-0.434	-0.436	0.055	7.908	0.000	**Strong negative;** higher fairness reduces perceived favouritism

All hypothesized paths were statistically significant ($p < 0.01$). The findings support partial mediation of Perceived Fairness in the relationship between Algorithmic Transparency and Workplace Favouritism, and a significant moderation effect of Digital Literacy on the PF → WF relationship. These results validate the model's theoretical assumptions and provide a strong quantitative foundation for the subsequent qualitative phase. These results validate the model's theoretical assumptions and provide a strong quantitative foundation for the subsequent qualitative phase.

6.6 Moderation Analysis

The moderating influence of DL on the relationship between PF and WF was further examined through an interaction plot generated in SmartPLS. The results revealed a clear negative moderation effect, as visualised in Fig. 5. The three regression lines correspond to different levels of DL: low (–1 SD), average (mean), and high (+ 1 SD). Each line represents how perceptions of fairness relate to favouritism at varying digital proficiency levels. The slope for employees with high DL (green line) is the steepest, indicating that an increase in PF leads to a substantial reduction in workplace favouritism. In contrast, for employees with low DL (red line), the slope is much flatter, suggesting that fairness perceptions have a weaker influence on reducing favouritism when employees lack the digital skills to interpret or evaluate algorithmic decisions accurately. The average group (blue line) falls between these two extremes,

representing a moderate decline in favouritism with increasing fairness perceptions. This pattern confirms the significant interaction effect identified in the bootstrapping results ($\beta = -0.208$, $t = 5.354$, $p < 0.001$), which demonstrates that DL moderates the fairness–favouritism relationship. Specifically, employees who possess higher digital competence are better able to comprehend the logic and fairness embedded in AI-driven HR systems, thereby reducing their perception of favouritism or bias. Conversely, those with lower digital literacy may lack confidence in algorithmic outcomes, making them more likely to perceive bias even in fair systems.

The moderation graph thus provides a visual validation of the statistical findings that DL strengthens the negative relationship between PF and WF. This result aligns with socio-technical, emphasising that technology's fairness outcomes are contingent upon user competence and understanding. In the context of AI-driven HR systems, higher DL not only enhances interpretability and trust but also enables employees to better recognise fairness cues, ultimately minimising PF in the workplace.

Ethics Statement

This research comprised two phases: a qualitative exploratory phase and a quantitative survey-based phase involving adult employees from the Indian IT sector. Neither phase involved clinical procedures, medical interventions, or vulnerable populations. In accordance with institutional research guidelines, anonymous and non-interventional social science research does not require formal ethics committee approval.

For both phases, participants were informed about the purpose of the study and the voluntary nature of their participation. Informed consent was obtained prior to data collection. Participants were informed of their right to withdraw at any time without consequence. No personally identifiable information was collected or stored in either phase, and all data were anonymized prior to analysis. All methods were performed in accordance with relevant institutional guidelines and regulations and adhered to internationally accepted ethical principles for research involving human participants, including the Declaration of Helsinki.

7. Discussions

The qualitative results offer deep contextual understanding of how employees understand AI-powered assessments, exposing issues that are consistent with current algorithmic management research. Participants consistently described the HR AI system as lacking clarity, context, and meaningful explanation reflecting what studies refer to as black-box opacity [30,35]. Concerns of this nature were linked to perceptions of mechanical fairness, in which employees acknowledged procedural equality yet believed that the system overlooked contextual subtleties and unrecognized contributions, reinforcing the notion that equal treatment does not ensure ethical fairness. Many employees reported a sense of algorithmic favouritism arising from data visibility bias, a phenomenon where roles with more quantifiable outputs receive unreasonable recognition, echoing insights from O'Neil [36] and Kroll et al. [73].

DL emerged as a powerful moderating factor in the lived experiences when those who understood how the system worked expressed less anxiety, greater trust, and reduced perceptions of bias, aligning with algorithmic literacy empowers users and mitigates fear [32,48]. Participants' emotional reactions ranging from initial confusion to eventual acceptance that demonstrate the psychological journey of employees when adapting to algorithmic systems, supporting previous findings that transparency and comprehension are essential for trust formation [3]. Collectively, the qualitative insights confirm that transparency, fairness, and literacy are not abstract constructs but deeply connected elements of employees' daily experiences in algorithmic workspaces.

The quantitative findings demonstrate how crucial algorithmic transparency is in influencing employees' views of fairness in AI-powered HR systems. Consistent with earlier research, transparency enhanced interpretability and reduced ambiguity surrounding algorithmic evaluations [42,43]. Employees were more inclined to see AI choices as procedurally fair if they believed they were intelligible, which supports justice theory viewpoints[33,34]. In line with socio-technical systems thinking, fairness perceptions emerged as the key psychological mechanism through which algorithmic design influences organisational outcomes, aligning with transparent AI strengthens perceived legitimacy and reduces perceived bias [37,74]. Furthermore, digital literacy strengthened the fairness–favouritism relationship, indicating that employees who were more technologically experienced will have fewer concerns about AI-driven partiality. This finding supports earlier work suggesting that algorithmic literacy reduce mistrust about bias to technological systems [32,75]. Taken together, the quantitative findings suggest that fairness perceptions are a crucial pathway through which transparency influences trust and reduces perceptions of favouritism in algorithmic HR environments. These findings also invite reflection on the broader organisational implications of placing AI systems within HR decision-making processes. When algorithmic evaluations are positioned as objective and neutral, they may inadvertently constrain employees' ability to question or contest outcomes, thereby reshaping traditional channels of voice and grievance. Research suggests that algorithmic authority often appears more legitimate precisely because it is perceived as technical rather than social, which can discourage critical engagement and normalize unequal outcomes [60]. In these situations, perceptions of bias can persist in subtle ways, not because of explicit bias but rather because there is a lack of clarity on the inclusion or exclusion of contextual aspects in algorithmic decisions.

The discussion further highlights that fairness perceptions in AI-mediated workplaces are not static but evolve through continuous social comparison and collective sensemaking. Employees interpret algorithmic outcomes not in isolation, but relative to peer experiences, informal performance norms, and shared narratives within teams. Sensemaking theory emphasises that individuals rely on social cues and retrospective interpretation to construct meaning around ambiguous events, particularly when technological systems obscure causal explanations. This perspective helps explain why transparency initiatives may initially enhance acceptance of AI systems but fail to fully eliminate concerns of favouritism if employees repeatedly observe misalignment between algorithmic outputs and everyday work realities.Another important implication concerns the shifting nature of power and responsibility in

algorithmic HR systems. While AI is often framed as a decision-support tool, employees may perceive managerial accountability as diluted when unfavorable outcomes are attributed to system logic. Socio-technical research warns that such displacement of responsibility can weaken trust and intensify perceptions of unfairness, particularly when employees are unsure who can be held accountable for correcting errors or biases [76]. This reinforces the importance of maintaining human oversight and clearly articulated governance structures alongside algorithmic implementation.

Finally, the findings suggest that the benefits of algorithmic transparency and digital literacy may not be evenly distributed across the workforce. While some employees are better equipped to interpret and navigate AI-driven evaluations, others may experience heightened vulnerability due to limited technological confidence or reduced access to explanatory resources. This uneven capacity for engagement echoes broader concerns about algorithmic inequality, where technological competence becomes an implicit criterion shaping perceptions of fairness and inclusion [77,78]. Consequently, organizations must recognize that fairness in algorithmic HR systems extends beyond technical design to include ongoing communication, support, and participatory practices that allow employees to meaningfully engage with AI-mediated decisions.

8.Implications

8.1 Theoretical Implications

This study contributes to theory by improving knowledge of how employees interpret algorithmic decision-making both emotionally and intellectually by extending organisational justice theory [33] and socio-technical systems theory [79] into the realm of AI-enabled HR practices. The confirmation that perceived fairness mediates the effect of transparency on favouritism strengthens theoretical models arguing that fairness is a psychological bridge between system design and user trust [37,42].The findings further enrich algorithmic management literature by demonstrating that data visibility bias produces new, non-human forms of favouritism that supporting critiques of algorithmic neutrality[36]. Additionally, the moderating role of digital literacy adds a novel theoretical dimension. Positioning literacy not merely as a skill but as a cognitive filter through which algorithmic fairness is interpreted extending research on algorithmic literacy and trust formation [32].

8.2 Practical Implications

Practically, the study highlights the urgent need for organisations to enhance the explainability of AI-driven HR systems. Through this, employees can interpret rating logic, understand data capture mechanisms, and connect their work to outputs. HR departments should introduce transparency dashboards, periodic explanation reports, and logic summaries faced by employee to reduce opacity and algorithmic anxiety [29]. Because digital literacy significantly reduces perceptions of favouritism. So organisations must implement structured AI literacy training, workshops, and user guides to enable employees to follow algorithmic tools with confidence. The qualitative results underscore the need to

integrate hybrid evaluation systems, where AI-generated scores are complemented by contextual input from managers to capture invisible labour such as mentoring, collaboration, and complex problem-solving. Implementing these would ensure that performance evaluation is not merely consistent but contextually informed and ethically grounded.

At the policy level, the findings support the development of institutional guidelines that mandate algorithmic transparency, fairness auditing, and employee participation in the design and deployment of HR technologies. Organisations should establish AI-based Governing Committees responsible for monitoring algorithmic drift, auditing fairness metric.Moreover Ensure that sensitive roles are not systematically disadvantaged due to visibility gaps [73]. Policy frameworks should require that employees receive access to model explanations and opportunities to clarify AI-generated evaluations. Additionally, national and organisational policies should promote digital literacy initiatives, ensuring usable access to algorithmic understanding across workforce segments. These policy measures will help to create AI ecosystems where transparency, fairness, accountability, and empowerment are embedded.

9. Limitations and Future Research Directions

This study provides significant insights into how algorithmic transparency, perceived fairness, digital literacy, and workplace favouritism interact in AI-driven HR systems. Some limitations must be acknowledged. First, the research was conducted within the Indian IT sector, which is highly digitised and technologically mature. Because of this, employee experiences in less technologically advanced businesses would vary, which would restrict how broadly the results can be applied. Second, data collection relied on self-reported perceptions in both survey and interview phases, which may be subject to social desirability bias, or subjective interpretation of algorithmic processes. Third, the qualitative component involved 20 participants, which, although sufficient for phenomenological saturation, may not fully capture variations across different organisational cultures and AI maturity levels. Additionally, the study examined digital literacy primarily at the individual level, whereas organisational digital readiness, infrastructure quality, and managerial support may also significantly shape fairness perceptions and trust. These contextual moderators were beyond the scope of the present work.

Future research can address these limitations by adopting cross-industry and cross-cultural comparisons to explore how different institutional and technological environments influence algorithmic fairness perceptions. Longitudinal studies would be particularly valuable in examining how trust, literacy, and fairness perceptions evolve as employees become more familiar with AI systems. Future researchers could also integrate experimental designs to isolate the causal effects of transparency interventions on fairness and favouritism perceptions. Additionally, there is scope for exploring team-level and organisational-level moderators such as digital culture, leadership support, and AI governance practices. As AI systems increasingly incorporate adaptive learning algorithms, future studies should investigate how algorithmic drift, data quality, and model updates shape user trust and perceived fairness over time. Finally, in order to better understand how employees deal with the challenges of

algorithmic management, future research should look at understudied characteristics including emotional resilience, resistance to automation, and the quality of human–AI collaboration.

10.Conclusion

According to the study's findings, the effectiveness and acceptance of AI-driven HR systems depend fundamentally on how employees interpret transparency, fairness, and their own digital capability within algorithmic workplaces. Together, the mixed-methods results demonstrate that algorithmic transparency plays a key role in influencing employees' assessments of fairness and validate that employees like systems that offer transparent data processing, comprehensive reasoning, and explicit explanations[42]. Qualitative insights added depth by showing that employees do not evaluate fairness passively; rather, they actively assess the interpretability of outcomes, contextual sensitivity of the system, and the ethical appropriateness of AI-based evaluations. Importantly, the qualitative narratives reveal that fairness is not merely the absence of bias but the presence of contextual understanding, explanation, and opportunities for employee involvement. While AI systems enhanced procedural consistency, they also introduced new forms of bias rooted in data visibility and context blindness, repeating debates in the algorithmic management literature regarding data-driven favouritism [36]. Quantitative results revealed that transparency enhances perceived fairness, which in turn significantly reduces perceptions of workplace favouritism, reinforcing long-standing justice theories that fairness is a critical psychological lens through which organisational decisions are evaluated [33].The study also highlights that digital literacy plays a transformative role, moderating the fairness–favouritism relationship by empowering employees to decode algorithmic outputs and interpret decisions more confidently [32]. Overall, the study concludes that building trust in AI-enabled HR systems requires more than technical accuracy. It requires explainability, participatory design, and digital empowerment to ensure that employees feel seen, understood, and fairly evaluated. By integrating transparency, fairness, and literacy into AI governance, organisations can reduce suspicion, strengthen trust, and promote unbiased human-centred algorithmic decision-making.

Declarations

Ethical Declaration

This study does not involve human subjects, animal experiments, or sensitive data. We strictly adhere to academic rigour and standards to ensure the scientific nature and reliability of the research

Competing interests

The authors declare no competing interests.

Author Details

Yajikumar E.M

Vellore Institute of Technology, Vellore, India

Sowmya G. (**Corresponding Author)**

Vellore Institute of Technology, Vellore, India, Email ID: sowmya.g@vit.ac.in

Funding

Acknowledgements

The authors received no financial support for the research, authorship, or publication of this article

Author Contribution

Conceptualisation, Methodology, Data collection, Formal analysis, writing original draft: Y.E.M.
Supervision, Writing review & editing: S.G

Acknowledgement

The authors would like to express sincere gratitude to all those who contributed, directly or indirectly, to the completion of this study. The authors also acknowledge the support received from the institution in providing the necessary academic environment and resources for conducting this research

Data Availability

Datasets used in this study are available from the corresponding author upon reasonable request.

References

1. Cappelli, P., Tambe, P. & Yakubovich, V. Artificial Intelligence in Human Resources Management: Challenges and a Path Forward. *SSRN J.* https://doi.org/10.2139/ssrn.3263878 (2018).
2. Meijerink, J., Boons, M., Keegan, A. & Marler, J. Algorithmic human resource management: Synthesizing developments and cross-disciplinary insights on digital HRM. *Int. J. Hum. Resource Manage.* **32**, 2545–2562 (2021).
3. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **3**, 2053951716679679 (2016).

4. Burrell, J. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big Data Soc.* **3**, 2053951715622512 (2016).
5. Soleimani, M., Arrowsmith, J., Intezari, A. & Pauleen, D. J. Mitigating bias in AI-powered HRM. in Research Handbook on Human Resource Management and Disruptive Technologies (eds (eds Bondarouk, T. & Meijerink, J.) 39–50 (Edward Elgar Publishing, doi:10.4337/9781802209242.00012. (2024).
6. Nguyen, D. S. W. & Shaik, M. M. Impact of Artificial Intelligence on Corporate Leadership. *JCC* 12, 40–48 (2024).
7. Narayanan, A. Understanding Social Media Recommendation Algorithms. https://doi.org/10.7916/KHDK-M460 doi:10.7916/KHDK-M460.
8. Slussareff, M. & O’Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. *CyberOrient* 16, 72–75 (2022). (2016).
9. Raji, I. D. et al. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. Preprint at (2020). https://doi.org/10.48550/ARXIV.2001.00973
10. Diakopoulos, N. Accountability in algorithmic decision making. *Commun. ACM*. **59**, 56–62 (2016).
11. Ananny, M. & Crawford, K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New. Media Soc.* **20**, 973–989 (2018).
12. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science* **356**, 183–186 (2017).
13. Binns, R. et al. ‘It’s Reducing a Human Being to a Percentage’; Perceptions of Justice in Algorithmic Decisions. (2018). https://doi.org/10.48550/ARXIV.1801.10408 doi:10.48550/ARXIV.1801.10408.
14. Lepri, B. et al. Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philos. Technol.* **31**, 611–627 (2018).
15. Kim, D. Y. & Kim, H. Y. Social media influencers as human brands: an interactive marketing perspective. *JRIM* **17**, 94–109 (2023).
16. Rupp, D. E., Shao, R., Jones, K. S. & Liao, H. The utility of a multifoci approach to the study of organizational justice: A meta-analytic investigation into the consideration of normative rules, moral accountability, bandwidth-fidelity, and social exchange. *Organ. Behav. Hum Decis. Process.* **123**, 159–185 (2014).
17. Colquitt, J. A. et al. Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives. *J. Appl. Psychol.* **98**, 199–236 (2013).
18. Shin, D. & Bianco, W. T. In Blockchain We Trust: Does Blockchain Itself Generate Trust? *Soc. Sci. Q.* **101**, 2522–2538 (2020).
19. Ouyang, W. Data Visualization in Big Data Analysis: Applications and Future Trends. *JCC* 12, 76–85 (2024).
20. Ng, W. Can we teach digital natives digital literacy? *Comput. Educ.* **59**, 1065–1078 (2012).

21. Eshet-Alkalai, Y. Thinking in the Digital Era: A Revised Model for Digital Literacy. *IISIT* 9, 267–276 (2012).
22. Van Deursen, A. J. & Van Dijk, J. A. The digital divide shifts to differences in usage. *New. Media Soc.* **16**, 507–526 (2014).
23. Khatri, N. & Tsang, E. W. K. Antecedents and Consequences of Cronyism in Organizations. *J. Bus. Ethics*. **43**, 289–303 (2003).
24. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* **54**, 1–35 (2022).
25. Raghavan, M., Barocas, S., Kleinberg, J. & Levy, K. Mitigating bias in algorithmic hiring: evaluating claims and practices. in *Proceedings of the Conference on Fairness, Accountability, and Transparency* 469–481 (ACM, Barcelona Spain, 2020). 469–481 (ACM, Barcelona Spain, 2020). (2020). 10.1145/3351095.3372828
26. Gordon, F. & Virginia Eubanks Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: Picador, St Martin's Press. *Law Tech Hum* 162–164 (2019) (2018). 10.5204/lthj.v1i0.1386
27. Creswell, J. W. & Plano Clark, V. L. *Designing and Conducting Mixed Methods Research* (Sage, 2018).
28. Venkatesh, V., Brown, S. A. & Bala, H. Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems1. *MIS Q.* **37**, 21–54 (2013).
29. Nicolas, M. L. D., Desroziers, A., Caccioli, F. & Aste, T. ESG reputation risk matters: An event study based on social media data. *Finance Res. Lett.* **59**, 104712 (2024).
30. Burrell, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **3**, 2053951715622512 (2016).
31. Tunstall, S. & Michigan State University. Models as Weapons: Review of Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy by Cathy O'Neil *Numeracy* 11, (2018). (2016).
32. Dogruel, L., Masur, P. & Joeckel, S. Development and Validation of an Algorithm Literacy Scale for Internet Users. *Communication Methods Measures*. **16**, 115–133 (2022).
33. Colquitt, J. A. On the dimensionality of organizational justice: A construct validation of a measure. *J. Appl. Psychol.* **86**, 386–400 (2001).
34. Cropanzano, R. S., Ambrose, M. L. & Bies, R. J. Interactional Justice. in The Oxford Handbook of Justice in the Workplace (eds (eds Cropanzano, R. S. & Ambrose, M. L.) (Oxford University Press, doi:10.1093/oxfordhb/9780199981410.013.4. (2015).
35. Selbst, A. D. & Barocas, S. The Intuitive Appeal of Explainable Machines. *SSRN J.* https://doi.org/10.2139/ssrn.3126971 (2018).
36. O'Neil, C. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (B/D/W/Y Broadway Books, 2017).

37. Lee, M. K. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data Soc.* **5**, 2053951718756684 (2018).
38. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. & Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **3**, 2053951716679679 (2016).
39. Wang, W., Siau, K., Artificial Intelligence, M. & Learning Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *J. Database Manage.* **30**, 61–79 (2019).
40. Long, D. & Magerko, B. What is AI Literacy? Competencies and Design Considerations. in Proceedings of the *CHI Conference on Human Factors in Computing Systems* 1–16 (ACM, Honolulu HI USA, 2020). HI USA, 2020). (2020). 10.1145/3313831.3376727
41. Dogruel, L., Masur, P. & Joeckel, S. Development and Validation of an Algorithm Literacy Scale for Internet Users. *Communication Methods Measures*. **16**, 115–133 (2022).
42. Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum. Comput. Stud.* **146**, 102551 (2021).
43. Nicolas, M. L. D., Desroziers, A., Caccioli, F. & Aste, T. ESG reputation risk matters: An event study based on social media data. *Finance Res. Lett.* **59**, 104712 (2024).
44. Rohden, S. F., Netto, S., Espartel, L. B. & C. F. & Autonomous technology in the marketplace: The impact of enjoyment on consumer responses. *Comput. Hum. Behav.* **168**, 108647 (2025).
45. Eiband, M. et al. ACM, Tokyo Japan,. Bringing Transparency Design into Practice. in *23rd International Conference on Intelligent User Interfaces* 211–223 (2018). 10.1145/3172944.3172961
46. Kizilcec, R. F. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. in *Proceedings of the CHI Conference on Human Factors in Computing Systems* 2390–2395 (ACM, San Jose California USA, 2016). 2390–2395 (ACM, San Jose California USA, 2016). (2016). 10.1145/2858036.2858402
47. Venkatesh, V., Thong, J. Y. L. & Xu, X. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology1. *MIS Q.* **36**, 157–178 (2012).
48. Mayer, R. E. How Multimedia Can Improve Learning and Instruction. in The Cambridge Handbook of Cognition and Education (eds (eds Dunlosky, J. & Rawson, K. A.) 460–479 (Cambridge University Press, doi:10.1017/9781108235631.019. (2019).
49. Alaimo, C., Kallinikos, J. & Valderrama, E. Platforms as service ecosystems: Lessons from social media. *J. Inform. Technol.* **35**, 25–48 (2020).
50. Greenberg, J. Organizational Justice: Yesterday, Today, and Tomorrow. *J. Manag.* **16**, 399–432 (1990).
51. Leventhal, G. S. What Should Be Done with Equity Theory? in Social Exchange (eds (eds Gergen, K. J., Greenberg, M. S. & Willis, R. H.) 27–55 (Springer US, Boston, MA, doi:10.1007/978-1-4613-3087-5_2. (1980).
52. Folger, R. & Cropanzano, R. Fairness theory: Justice as accountability. in 1–55 (2001).
53. Prendergast, C. & Topel, R. H. Favoritism in Organizations. *J. Polit. Econ.* **104**, 958–978 (1996).

54. Ambrose, M. L. & Schminke, M. The role of overall justice judgments in organizational justice research: A test of mediation. *J. Appl. Psychol.* **94**, 491–500 (2009).
55. Thibaut, J. & Walker, L. A Theory of Procedure. *Calif. Law Rev.* **66**, 541 (1978).
56. Van Den Broeck, A., Howard, J. L., Van Vaerenbergh, Y., Leroy, H. & Gagné, M. Beyond intrinsic and extrinsic motivation: A meta-analysis on self-determination theory's multidimensional conceptualization of work motivation. *Organizational Psychol. Rev.* **11**, 240–273 (2021).
57. Miles, M. B. & Huberman, A. M. *Qualitative Data Analysis: An Expanded Sourcebook*.
58. Tashakkori, A., Johnson, R. B. & Teddlie, C. *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences* (SAGE, 2021).
59. Ribeiro, M. T., Singh, S. & Guestrin, C. 'Why Should I Trust You?': Explaining the Predictions of Any Classifier. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1135–1144ACM, San Francisco California USA, (2016). 10.1145/2939672.2939778
60. Kellogg, K. C., Valentine, M. A. & Christin, A. Algorithms at Work: The New Contested Terrain of Control. *ANNALS* **14**, 366–410 (2020).
61. Meijerink, J. & Bondarouk, T. The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. *Hum. Resource Manage. Rev.* **33**, 100876 (2023).
62. Wenzel, M., Stanske, S. & Marvin, B. Lieberman. Strategic responses to crisis. *Strategic Manage. Journal* **42**, (2021).
63. Cropanzano, R., Anthony, E. L., Daniels, S. R. & Hall, A. V. Social Exchange Theory: A Critical Review with Theoretical Remedies. *ANNALS* 11, 479–516 (2017).
64. Khatri, N. & Tsang, E. W. K. Antecedents and Consequences of Cronyism in Organizations. *J. Bus. Ethics*. **43**, 289–303 (2003).
65. Shin, D. & Park, Y. J. Role of fairness, accountability, and transparency in algorithmic affordance. *Comput. Hum. Behav.* **98**, 277–284 (2019).
66. Colom, A. & van Dijk, C. The Digital Divide: by Jan Polity Press, 208 pp., £17.99 (paperback), ISBN: 978-1-509-534456. *Information, Communication & Society* 23, 1706–1708 (2020). (2020).
67. Cheng, M. M. & Hackett, R. D. A critical review of algorithms in HRM: Definition, theory, and practice. *Hum. Resource Manage. Rev.* **31**, 100698 (2021).
68. Lepri, B. et al. Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philos. Technol.* **31**, 611–627 (2018).
69. Straub, D. & Gefen, D. Validation Guidelines for IS Positivist Research. *CAIS* 13, (2004).
70. Hair, J. F. *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook* (Springer International Publishing AG, 2021).
71. Fornell, C. & Larcker, D. F. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *J. Mark. Res.* **18**, 39 (1981).
72. Gignac, G. E. Psychometrics and the Measurement of Emotional Intelligence. in Assessing Emotional Intelligence (eds (eds Parker, J. D. A., Saklofske, D. H. & Stough, C.) 9–40 (Springer US,

Boston, MA, doi:10.1007/978-0-387-88370-0_2. (2009).

73. Joshua, A. Kroll & Joanna Huey. Accountable Algorithms. *University Pa. Law Review* **165**, (2017).
74. Wang, W., Siau, K., Artificial Intelligence, M. & Learning Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *J. Database Manage.* **30**, 61–79 (2019).
75. Long, D. & Magerko, B. What is AI Literacy? Competencies and Design Considerations. in Proceedings of the *CHI Conference on Human Factors in Computing Systems* 1–16 (ACM, Honolulu HI USA, 2020). HI USA, 2020). (2020). 10.1145/3313831.3376727
76. Orlikowski, W. J. & Scott, S. V. 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization. *ANNALS* 2, 433–474 (2008).
77. D'Ignazio, C. & Klein, L. F. *Data Feminism* (The MIT Press, 2020). 10.7551/mitpress/11805.001.0001
78. Gran, A. B., Booth, P. & Bucher, T. To be or not to be algorithm aware: a question of a new digital divide? *Inform. Communication Soc.* **24**, 1779–1796 (2021).
79. Bostrom, N. The Vulnerable World Hypothesis. *Global Policy*. **10**, 455–476 (2019).

Figures

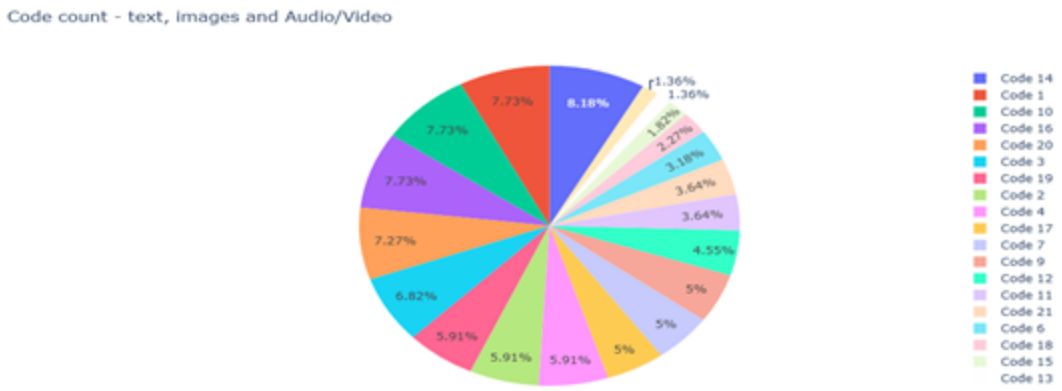


Figure 1

Code Count Pie Chart

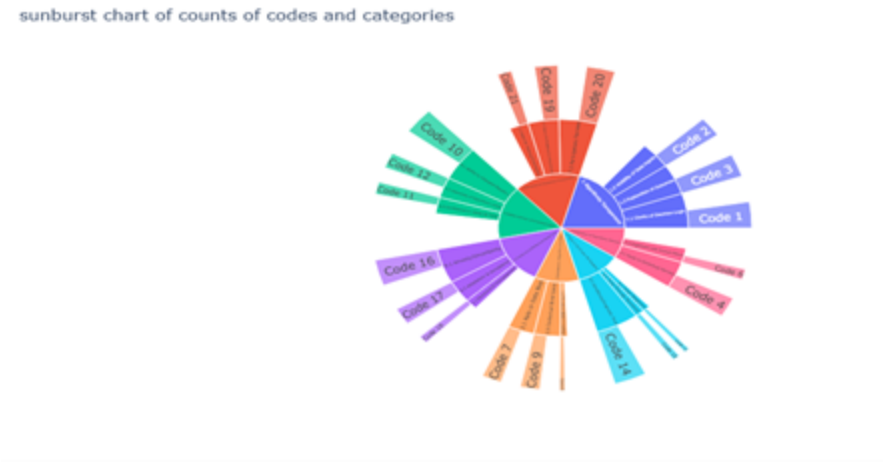


Figure 2

Sunburst Diagram

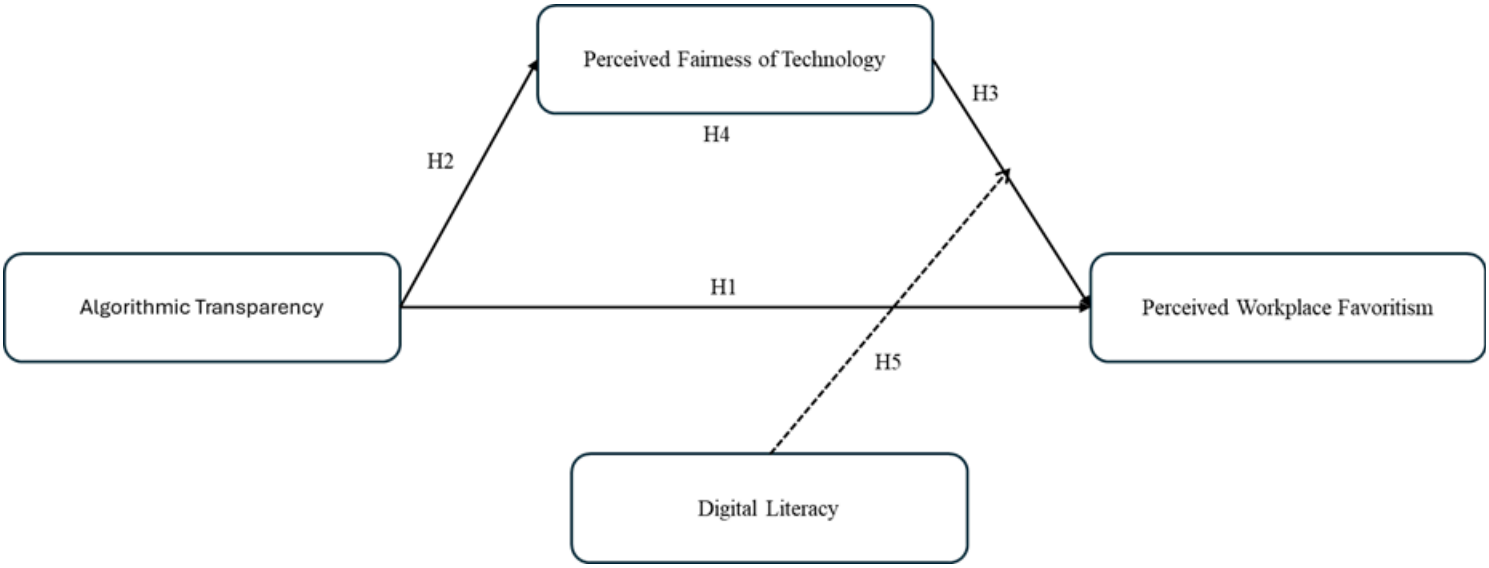


Figure 3

Conceptual framework

Source: Authors' own work

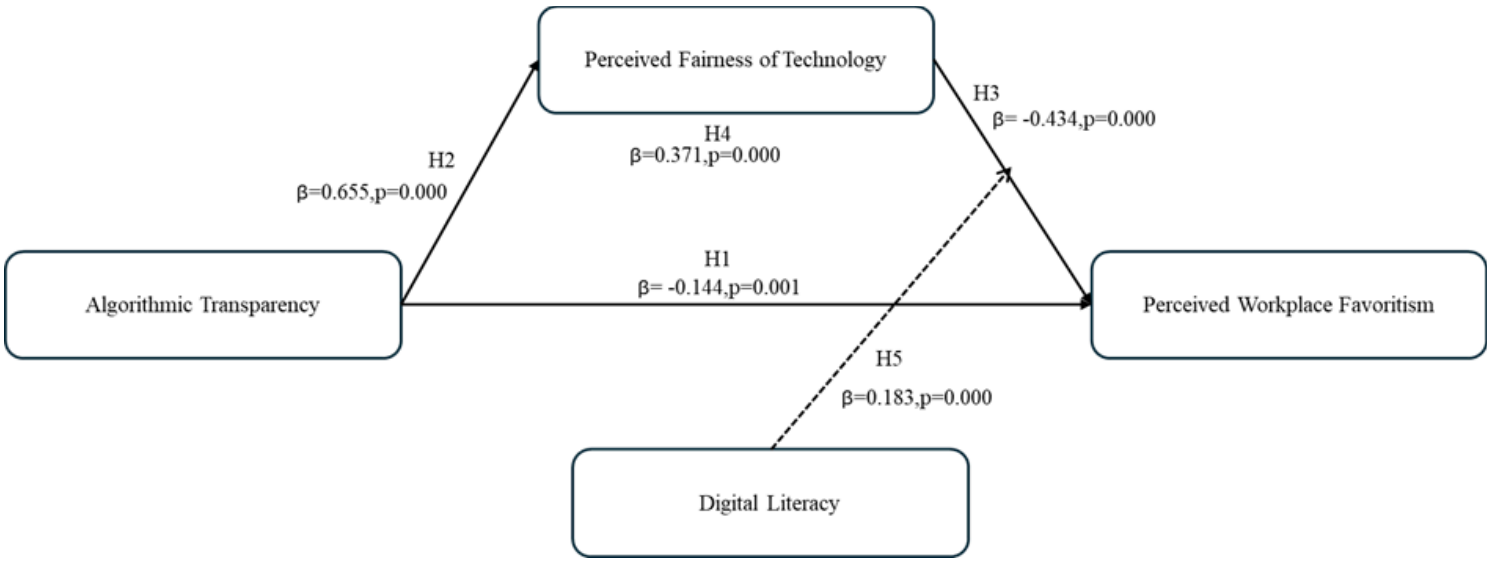


Figure 4

Structural model estimates

(Source: Author's own work).

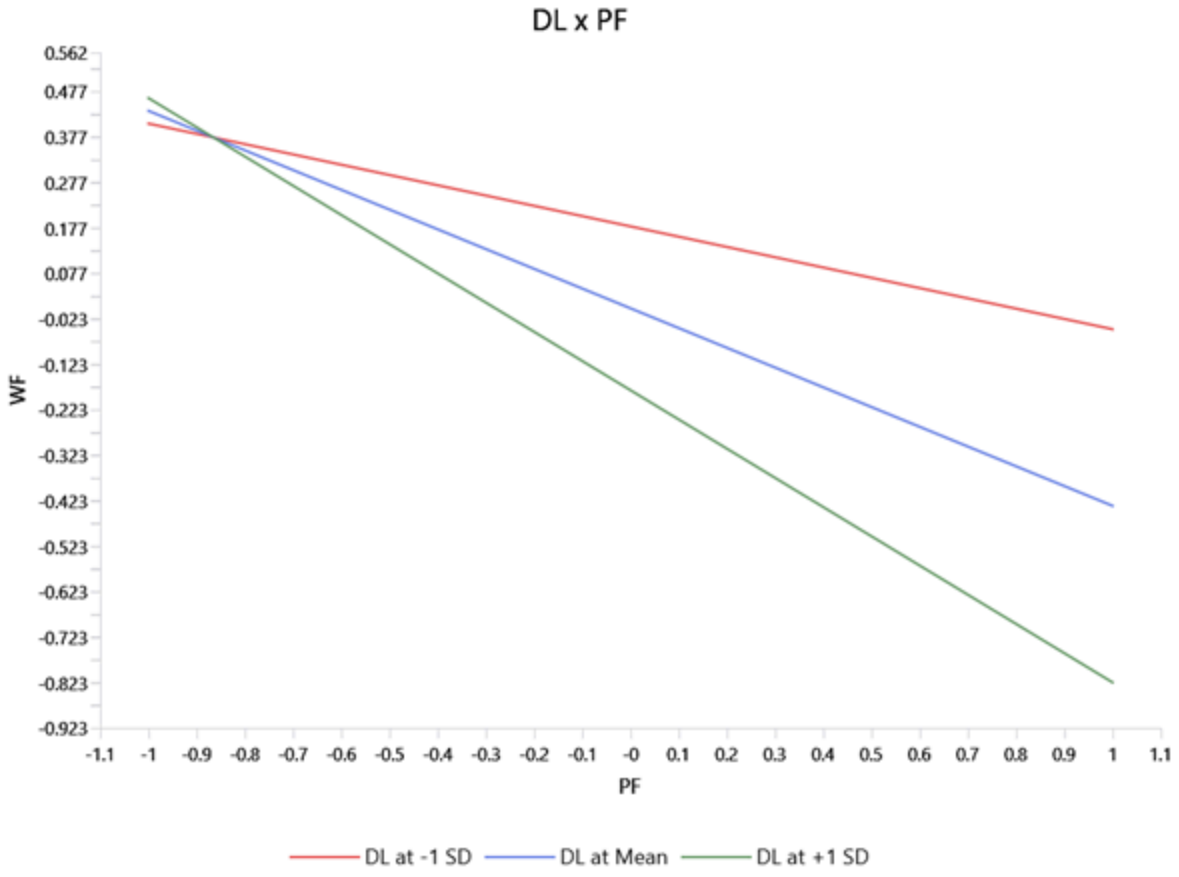


Figure 5

Figure 1: Moderation Effect of Digital Literacy on the Fairness–Favouritism Relationship