International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com OPEN ACCESS E-ISSN : 3049-2424
Volume 3 Issue 7 , July 2026 DOI: https://doi.org/10.71366/ijwos03072671072

Employee Perceptions of AI Fairness in Software Development Organizations:Toward an Integrative Conceptual Model of Algorithmic Trust and Organizational Justice

M. Farina Begam[1], Dr. S.Sudhamathi[2]
[1] *Research Scholar, Alagappa Institute of management, Alagappa University*
[2] *Associate Professor, Alagappa Insitute of Management, Alagappa University.*

Article Info

Article History:

Published: 16 July 2026

Publication Issue:
Volume 3, Issue 7
July-2026

Page Number:
289-296

Corresponding Author:
M. Farina Begam

Abstract:

Software organizations are among the most intensive adopters of workplace artificial intelligence (AI), embedding intelligent systems directly inside technical work through AI-assisted code generation, automated code review, agentic coding assistants, and algorithmic productivity dashboards. A substantial body of research has examined how employees perceive the fairness of AI-based decisions in administrative human-resource contexts such as hiring, performance appraisal, and career development, and this literature consistently shows that fairness perceptions, not technical accuracy alone, determine whether employees accept, resist, or feel alienated by algorithmic decisions. However, comparatively little conceptual work has addressed how software developers specifically perceive the fairness of AI systems embedded in their core technical craft. This paper synthesizes organizational justice theory, algorithmic fairness research, and the software-engineering trust literature to identify four interrelated gaps: the concentration of fairness research in administrative rather than technical knowledge-work contexts, the inconsistent conceptualization of trust in software engineering research, insufficient attention to contextual and experience-level variation among developers, and the absence of an integrative model linking organizational justice to algorithmic trust and professional outcomes in technical settings. Building on Colquitt's four-dimensional justice framework, this paper proposes an integrative model in which Perceived AI Fairness in Technical Work, comprising distributive, procedural, interpersonal, and informational dimensions, shapes technical and relational algorithmic trust, which in turn influences developer engagement, professional identity, knowledge-sharing behavior, and turnover intention. AI transparency, organizational communication climate, developer experience level, and ethical leadership are proposed as boundary conditions. Six propositions are advanced to guide future empirical validation, including quantitative survey designs suitable for PLS-SEM analysis. The paper contributes a theoretically grounded and testable framework for understanding fairness perceptions in AI-augmented technical work, and offers practical guidance for software organizations seeking to implement AI in ways developers experience as fair and trustworthy.

Keywords: employee perceptions, artificial intelligence, organizational justice, algorithmic trust, software development, ethical AI, conceptual model

1. Introduction

Software development organizations have become one of the earliest and most intensive testbeds for workplace AI. Unlike industries where AI mainly supports discrete administrative decisions, software organizations increasingly embed AI directly inside the technical work itself: AI pair-programming assistants generate and refactor code, automated systems triage and comment on pull requests, and agentic coding assistants now plan and execute multi-

step technical work with meaningful autonomy rather than simply completing lines of code (Roychoudhury et al., 2026). This depth of integration means that AI in software organizations is not a peripheral tool but an active participant in the very activity, writing and evaluating code, through which developers construct their professional identity and expertise.

Running parallel to this technical diffusion, a substantial and rapidly growing research stream has examined how employees and job applicants perceive AI-based decisions in domains such as personnel selection, performance evaluation, and career development. This literature consistently finds that fairness perceptions, rather than technical accuracy alone, determine whether people accept, resist, or feel alienated by algorithmic decisions (Malin et al., 2025; Qin et al., 2023; Köchling et al., 2024). Explanations and transparency about how an AI system reached a decision have repeatedly been shown to increase perceived fairness, sometimes bringing evaluations of AI-based decisions up to, or even above, the level of human decisions made without explanation (Malin et al., 2025). Yet this body of work is concentrated overwhelmingly in administrative human-resource contexts: hiring, résumé screening, and performance appraisal. Comparatively little conceptual attention has been directed at how technical professionals, and software developers in particular, perceive the fairness of AI systems that are woven into their day-to-day craft.

This omission matters because trust and fairness judgments in technical, creative knowledge work are entangled with professional identity, craft ownership, and epistemic authority in ways that differ from typical administrative-decision contexts. Recent software-engineering research has begun to note that the concept of trust itself is used inconsistently in this literature, often reduced informally to a willingness to accept AI-generated content rather than grounded in established psychological models of trust (Baltes et al., 2026). At the same time, practitioner and industry commentary has raised concerns that overreliance on AI-generated code, particularly among less experienced developers, may erode the very skills and judgment that fairness and quality assurance in software work depend on. These threads, organizational justice, algorithmic fairness, and software-engineering trust, have developed largely in parallel and have rarely been brought into conversation with one another.

This paper addresses that gap by developing a conceptual model of software developers' perceptions of AI fairness. It grounds the model in organizational justice theory (Colquitt, 2001), extends it into the technical work context, and integrates it with the software-engineering trust literature's recent call for more rigorous conceptualization of trust in AI-assisted technical work (Baltes et al., 2026). The paper's purpose is threefold: first, to synthesize literatures that have rarely been considered together; second, to identify specific, actionable research gaps; and third, to propose an integrative conceptual model with testable propositions capable of guiding future empirical, quantitative research, including survey-based designs suitable for partial least squares structural equation modeling (PLS-SEM).

2. Theoretical Background

2.1 AI Adoption in Software Development Organizations

AI now touches nearly every stage of the software development lifecycle. Code-completion and code-generation tools assist with routine and increasingly complex implementation tasks; automated review tools scan pull requests for

defects, security anti-patterns, and style violations; and agentic coding assistants can independently plan a body of work, invoke development tools, and iterate on a solution with only limited human direction (Roychoudhury et al., 2026). This growing autonomy raises the stakes considerably: whereas earlier tools offered static suggestions that a developer could accept or reject, agentic systems now make a stream of decisions on the developer's behalf, decisions whose fairness and trustworthiness the developer typically discovers only after the fact, through code review. Industry commentary has already flagged a related risk: as review shifts from evaluating human-authored code to evaluating AI-authored code, the perceived fairness and adequacy of that review process itself becomes a live organizational concern, particularly when AI-generated code is reviewed by another AI system with limited human oversight.

2.2 From Technology Acceptance to Fairness Perceptions

Historically dominant technology-adoption frameworks such as the Unified Theory of Acceptance and Use of Technology (UTAUT; Venkatesh et al., 2003) explain adoption largely through performance expectancy, effort expectancy, and social influence. These models remain valuable for explaining whether employees intend to use a new AI tool, and they underpin readiness-for-change research more broadly (Armenakis et al., 1993). However, once AI outputs begin to shape evaluative, reputational, or career-relevant outcomes, acceptance-oriented models say relatively little about the ethical and relational dimensions of the interaction: whether the process felt fair, whether the employee felt respected, and whether the reasoning behind an AI's output was adequately explained. The employee-perception literature on AI has accordingly shifted much of its theoretical weight toward organizational-justice-based fairness constructs, which better capture these relational and ethical dimensions (Qin et al., 2023; Köchling et al., 2024).

2.3 Organizational Justice Theory

Organizational justice theory offers the most well-validated framework for understanding fairness perceptions at work. Colquitt's (2001) four-factor model distinguishes distributive justice (the perceived fairness of outcomes), procedural justice (the perceived fairness of the processes used to reach those outcomes), interpersonal justice (the degree of respect and dignity with which people are treated), and informational justice (the adequacy and honesty of the explanations provided). This four-factor structure has consistently outperformed two- and three-factor alternatives, and each dimension carries distinct predictive value for outcomes such as commitment, rule compliance, and satisfaction (Colquitt, 2001). The framework has since been extended productively into AI-related decision contexts: Gilliland's (1993, 2001) foundational work on applicant reactions to selection systems has been repeatedly adapted to AI-based hiring and rejection scenarios, and recent research demonstrates that Colquitt's justice constructs, originally designed for human-administered processes, remain applicable, with some adaptation, to algorithmic decision-making (Ochmann et al., 2024; Malin et al., 2025).

2.4 Algorithmic Trust in Software Engineering

Trust is central to whether developers adopt and appropriately rely on AI-assisted tools, but software-engineering research has treated the construct inconsistently. A recent critical review found that trust is rarely explicitly defined in software-engineering articles on AI assistants, and is often conflated with the narrower notion of content-acceptance,

that is, whether a developer accepts a given AI suggestion, rather than being grounded in established psychological trust models (Baltes et al., 2026). Complementary work distinguishes a technical dimension of trust, developers' confidence that AI-generated code is correct, secure, and maintainable, from a more relational, human dimension of trust that depends on explainability, alignment with team norms and values, and the AI system's ability to adapt to feedback (Roychoudhury et al., 2026). This distinction is conceptually close to, yet has not been connected with, the interpersonal and informational dimensions of organizational justice theory, suggesting a natural point of theoretical integration.

3. Identifying the Research Gap

Bringing together the literatures reviewed above surfaces four interrelated and, to date, largely unaddressed gaps.

Gap 1: Contextual concentration in administrative decisions

Fairness-perception research on workplace AI is concentrated almost entirely in administrative human-resource contexts, personnel selection, performance evaluation, and career development (Malin et al., 2025; Qin et al., 2023; Köchling et al., 2024). None of this literature addresses the distinctly technical, creative, and identity-laden context of software engineering, in which the object of AI-mediated judgment is often the developer's own code, and in which the developer is simultaneously a user of AI, a reviewer of AI output, and a professional whose expertise is implicitly being compared against the machine's.

Gap 2: Fragmented conceptualization of trust

Software-engineering research on AI assistants rarely defines trust rigorously, frequently reducing it to willingness to accept AI-generated content (Baltes et al., 2026). This fragmentation makes it difficult to relate software-engineering findings on trust to the well-validated organizational justice literature, and prevents cumulative theory-building across the two fields.

Gap 3: Unexamined contextual and cross-cultural variation

Scholars have explicitly noted that fairness perceptions of AI vary across social, organizational, and cultural contexts (Hunkenschroer & Luetge, 2022), yet the organizational cultures distinctive to software development, open-source versus proprietary codebases, agile team structures, and distributed or remote engineering teams, remain unexamined through an organizational-justice lens.

Gap 4: Absence of an integrative model with technical, professional outcomes

No existing framework connects organizational justice dimensions to software-engineering-specific trust constructs and to the professional and psychological outcomes distinctive to developers: craft ownership, professional identity, skill development concerns, and knowledge-sharing behavior. Existing fairness research in AI-augmented performance evaluation shows that employees' fairness judgments shape engagement and career-relevant behaviors in general organizational settings (Qin et al., 2023), but this evidence has not been extended to the specific mechanisms of technical, code-centered work, nor has developer experience level been formally examined as a boundary condition despite preliminary industry evidence that junior and senior developers relate to AI-generated work quite differently.

Taken together, these gaps create a clear opportunity for a conceptual model that (a) relocates organizational justice theory into the software-engineering technical context, (b) integrates it with a more rigorously specified conceptualization of algorithmic trust, and (c) specifies professional-identity-relevant outcomes and boundary conditions specific to developers.

4. Toward an Integrative Conceptual Model

Building on the preceding synthesis, this paper proposes an integrative construct, Perceived AI Fairness in Technical Work (PAF-TW), adapted from Colquitt's (2001) four justice dimensions to the specific setting of AI-augmented software development. Table 1 summarizes the model's core constructs.

Construct	Dimension	Description in the technical work context
Perceived AI Fairness in Technical Work (PAF-TW)	Distributive fairness	Perceived equity of AI-mediated outcomes relative to effort and contribution (e.g., which code suggestions are merged, how productivity metrics are computed).
	Procedural fairness	Perceived neutrality, consistency, and transparency of the process by which AI systems reach suggestions or decisions (e.g., how a review bot flags issues, how an agent plans its workflow).
	Interpersonal fairness	Perceived respect and consideration in how the organization deploys AI, including whether developers are consulted rather than having tools imposed on them.
	Informational fairness	Adequacy and honesty of explanations for why an AI system produced a particular output or flagged a particular issue (explainability).
Algorithmic Trust	Technical trust	Confidence in the correctness, security, and maintainability of AI-generated outputs.
	Relational trust	Confidence that the AI-mediated process respects developer expertise, autonomy, and values.

Professional Outcomes	Identity, engagement, knowledge-sharing, turnover intention	Downstream developer outcomes hypothesized to follow from algorithmic trust.
Boundary Conditions	Transparency, communication climate, experience level, ethical leadership	Moderators expected to strengthen or weaken the fairness-trust-outcome pathway.

Table 1. Core constructs of the proposed integrative conceptual model.

In the proposed model, the four PAF-TW dimensions are hypothesized to shape algorithmic trust, operationalized as the two related but distinct dimensions of technical trust and relational trust identified in the software-engineering literature (Roychoudhury et al., 2026). Algorithmic trust, in turn, is hypothesized to influence a set of professional outcomes distinctive to technical knowledge work: professional identity and craft ownership (whether developers feel their expertise still matters), work engagement and job satisfaction, knowledge-sharing and collaborative verification behavior (willingness to mentor others and share what has been learned about an AI tool's blind spots), and turnover intention. Four boundary conditions are proposed to moderate these relationships: the transparency and explainability designed into AI tools, the organization's communication climate around AI (echoing the training-and-communication emphasis found in readiness-for-change research), developer experience level, and the ethical leadership exercised by engineering managers, which prior work links to organizational justice perceptions and, in turn, to ethical workplace behavior more broadly.

5. Propositions

Proposition 1: Higher perceived distributive fairness of AI-mediated technical outcomes is positively associated with developers' technical trust in AI systems.

Proposition 2: Higher perceived procedural fairness of AI decision-making processes is positively associated with developers' relational trust in AI systems.

Proposition 3: Interpersonal and informational fairness perceptions are positively associated with developers' willingness to engage in knowledge-sharing and collaborative verification of AI-generated code.

Proposition 4: Algorithmic trust (technical and relational) mediates the relationship between PAF-TW dimensions and developer outcomes, including engagement, professional identity, and turnover intention.

Proposition 5: AI transparency and explainability strengthen the positive relationship between procedural and informational fairness and algorithmic trust.

Proposition 6: Developer experience level moderates the fairness-trust relationship, such that junior developers exhibit a more reliance-driven trust response while senior developers exhibit a more scrutiny-driven, justice-sensitive trust response.

6. Discussion and Theoretical Contributions

This conceptual model makes three principal contributions. First, it extends organizational justice theory into a technical work context in which the object of judgment is the employee's own creative and technical output, rather than an administrative decision made about the employee, a meaningfully different psychological situation that existing justice-based AI research has not addressed. Second, it responds directly to the software-engineering literature's own call for a more rigorously grounded conceptualization of trust (Baltes et al., 2026) by anchoring algorithmic trust in an established, multi-dimensional justice framework rather than treating trust as synonymous with content acceptance. Third, it offers a bridge between two literatures, human-resource fairness research and software-engineering trust research, that have developed largely in isolation from one another despite addressing closely related psychological phenomena.

7. Practical Implications

For software organizations, the model suggests several concrete design and management priorities. Building explainability into AI coding tools and productivity dashboards addresses informational fairness directly. Involving developers in decisions about which AI tools are adopted and how they are governed addresses procedural and interpersonal fairness. Differentiating onboarding and training by experience level acknowledges the proposed moderating role of career stage, with junior developers likely needing more structured guidance on when to scrutinize AI output rather than defer to it. Finally, investing in transparent, consistent communication about how and why AI tools are being introduced, and in engineering leadership that models ethical, fairness-conscious use of AI, are practical levers directly implied by the model's boundary conditions.

8. Limitations and Directions for Future Research

As a conceptual paper, this model's propositions require empirical validation before they can inform practice with confidence. A natural next step is a cross-sectional quantitative survey of technical staff at AI-adopting software organizations, measuring the PAF-TW dimensions, technical and relational trust, and the proposed outcomes, analyzed using PLS-SEM given the model's multiple mediated and moderated pathways and its suitability for exploratory, theory-building research with modest sample sizes. Complementary qualitative work would help refine PAF-TW measurement items specific to software engineering, since existing justice scales were developed for administrative decision contexts. Longitudinal designs would be valuable for examining whether skill-erosion concerns raised in the practitioner literature materialize over time, and cross-cultural replication would address the contextual variation gap identified by Hunkenschroer and Luetge (2022). Future work might also examine how this fairness-trust model interacts with more traditional technology-acceptance variables such as those captured by UTAUT (Venkatesh et al., 2003), to clarify whether fairness perceptions operate as an antecedent to, or independently of, adoption intentions.

9. Conclusion

As AI becomes an active participant in software engineering rather than a peripheral administrative tool, understanding how developers judge the fairness of AI-mediated technical decisions becomes a pressing theoretical and practical concern. This paper has argued that existing fairness research, concentrated in administrative human-resource

contexts, and existing software-engineering trust research, often undertheorized, have each captured only part of this picture. By proposing an integrative model of Perceived AI Fairness in Technical Work, grounded in organizational justice theory and connected to software-engineering-specific conceptions of trust, this paper offers a testable framework for future empirical research and a set of practical levers for organizations seeking to implement AI in ways their technical employees experience as fair, transparent, and respectful of their professional expertise.

References

1. Armenakis, A. A., Harris, S. G., & Mossholder, K. W. (1993). Creating readiness for organizational change. Human Relations, 46(6), 681-703.
2. Baltes, S., Speith, T., Chiteri, B., Mohsenimofidi, S., Chakraborty, S., & Buschek, D. (2026). On the need to rethink trust in AI assistants for software development: A critical review. IEEE Transactions on Software Engineering.
3. Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. Journal of Applied Psychology, 86(3), 386-400.
4. Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. Journal of Applied Psychology, 86(3), 425-445.
5. Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. Academy of Management Review, 18(4), 694-734.
6. Gilliland, S. W., Groth, M., Baker, R. C., Dew, A. F., Polly, L. M., & Langdon, J. C. (2001). Improving applicants' reactions to rejection letters: An application of fairness theory. Personnel Psychology, 54(3), 669-703.
7. Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. Journal of Business Ethics, 178(4), 977-1007.
8. Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research, 13, 795-848.
9. Köchling, A., Wehner, M. C., & Ruhle, S. A. (2024). This (AI)n't fair? Employee reactions to artificial intelligence (AI) in career development systems. Review of Managerial Science, 19(4), 1195-1228.
10. Malin, C., Fleiß, J., Ortlieb, R., & Thalmann, S. (2025). Rejected by an AI? Comparing job applicants' fairness perceptions of artificial intelligence and humans in personnel selection. Frontiers in Artificial Intelligence, 8, Article 1671997.
11. Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Lammer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. Information Systems Journal, 34(2), 384-414.
12. Qin, S. (M.), Jia, N., Luo, X., Liao, C., & Huang, Z. (2023). Perceived fairness of human managers compared with artificial intelligence in employee performance evaluation. Journal of Management Information Systems, 40(4), 1039-1070.
13. Roychoudhury, A., Păsăreanu, C., Pradel, M., & Ray, B. (2026). Agentic AI software engineers: Programming with trust. Communications of the ACM, 69(5), 56-58.
14. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478.