E-ISSN: 3108-4192

APSSHS

Academic Publications of Social Sciences and Humanities Studies
2024, Volume 4, Issue 2, Page No: 103-111
Available online at: https://apsshs.com/

Journal of Applied Organizational Systems and Behavior

Algorithmic Psychological Contracts and Fairness Expectations in Artificial Intelligence–Mediated Work

Olivia Bennett[1]*, Harry Collins[2], Jack Foster[1], Ethan Wright[3]

1. Department of Algorithmic Psychological Contracts, Faculty of Business, University of Nottingham, Nottingham, United Kingdom.
2. Department of Fairness and AI-Mediated Work, Faculty of Management, University of Southampton, Southampton, United Kingdom.
3. Department of Work Expectations, Faculty of Psychology, Newcastle University, Newcastle, United Kingdom.

Abstract

Artificial intelligence increasingly mediates task allocation, performance evaluation, scheduling, selection, and other consequential employment decisions. Yet organizational theory has not fully explained how employees form reciprocal expectations when managerial claims are communicated or enacted through algorithmic systems rather than through direct human interaction. This original conceptual theory article develops Algorithmic Psychological-Contract Theory to explain how organizational claims, prior exchange schemas, system roles, and enacted algorithmic practices may become cues from which employees infer obligations. The theory distinguishes general expectations, fairness judgments, trust, formal policy, and technological functionality from perceived reciprocal obligations attributed to the organization. It proposes that consequential algorithm-mediated encounters activate discrepancy, attribution, fairness, and legitimacy appraisals, producing perceived fulfilment, ambiguity, or breach. These appraisals may subsequently shape emotion, trust recalibration, voice, silence, resistance, withdrawal, and contract updating. Transparency is treated not as disclosure alone but as a possible organizational claim whose promissory force depends on comprehensibility, accountability, contestability, and consequential review. The article also proposes procedural, relational, structural, and participatory repair pathways, while emphasizing that their adequacy should depend on the inferred obligation and perceived source of breach. The contribution is a bounded socio-technical theory that retains organizational accountability without treating algorithms as autonomous contracting parties. Its relationships and propositions remain conceptually plausible rather than empirically validated and require discriminant measurement, longitudinal process research, field experimentation, qualitative tracing, and multilevel tests across technologies, occupations, employment arrangements, and institutional settings.

Keywords: Psychological contracts beyond human managers, Conceptual foundations, Fairness expectations toward algorithmic systems, Transparency claims and inferred obligations, Algorithmic psychological-contract theory, Propositions and repair pathways

How to cite this article: Bennett O, Collins H, Foster J, Wright E. Algorithmic Psychological Contracts and Fairness Expectations in Artificial Intelligence–Mediated Work. J Appl Organ Syst Behav. 2024;4(2):103-11. https://doi.org/10.51847/o95HH2b6Sl

Received: 15 March 2024; **Revised:** 24 June 2024; **Accepted:** 27 June 2024
Corresponding author: Olivia Bennett
E-mail ✉ olivia.bennett@nottingham.ac.uk

Introduction

Psychological contracts concern employees' perceptions of reciprocal obligations within an exchange relationship, rather than every hope, preference, expectation, or judgment about organizational treatment. Their analytical value lies in explaining how employees interpret what the organization owes them, what they owe in return, and whether those perceived obligations have been fulfilled or breached [1]. This distinction becomes increasingly important when employment cues are delivered through

systems that appear impersonal yet materially shape access to work, evaluation, reward, development, and continued participation.

Algorithmic management changes how workers receive, interpret, and respond to organizational exchange cues. Rules embedded in interfaces, automated recommendations, scores, rankings, notifications, and workflow constraints may communicate what performance is valued, what treatment is promised, and what forms of reciprocity are expected [2]. However, an algorithmic output is not automatically a promise. Its potential contractual significance depends on whether employees attribute the cue to the organization, understand it as obligation-bearing, and connect it to an exchange relationship rather than merely to technical operation.

Algorithms can redistribute workplace control by directing activity, evaluating performance, and disciplining conduct while also creating new sites of worker interpretation and contestation [3]. This redistribution matters because a consequential decision may be presented as technologically determined even when the organization selected the data, objectives, thresholds, workflow, and authority structure. The visible decision source and the accountable exchange partner may therefore diverge. A theory centered only on human managerial statements misses how organizational treatment is increasingly enacted through socio-technical arrangements.

Artificial intelligence can automate some judgment while augmenting human decision-making elsewhere, making agency and responsibility contingent rather than binary [4]. Employees may encounter a manager who relies on a recommendation, a system that limits managerial discretion, or a workflow in which responsibility is dispersed among designers, vendors, analysts, and line managers. The relevant theoretical question is consequently not whether the human or the machine "made" the decision, but how employees interpret organizational authorship, authorization, and responsibility across the arrangement.

Work design further complicates attribution. Automation and algorithms can alter autonomy, task variety, feedback, workload, monitoring, role clarity, and responsibility, with potentially enabling or constraining consequences depending on design and context [5]. Reactions attributed to unfair algorithms may therefore partly reflect redesigned work, intensified demands, reduced discretion, or insecurity. Any theory of algorithm-mediated obligations must treat these conditions as alternative explanations and boundary conditions rather than assuming that every adverse response reflects contractual breach.

This article develops an original, non-empirical Algorithmic Psychological-Contract Theory. It proposes that algorithms can transmit and enact organizational claims without becoming independent contracting parties; that fairness and transparency become possible contract content only through reciprocity-based inference; and that breach requires an inferred obligation, a consequential discrepancy, and organizational attribution. Algorithmic systems redistribute managerial control, agency, and employment cues while remaining embedded in organizational arrangements [2-4]. The contribution is therefore a bounded account of formation, fulfilment, breach, response, and repair that distinguishes conceptual plausibility from empirical validation.

Conceptual foundations

The proposed theory begins with social exchange but does not equate social exchange, organizational justice, trust, and psychological contracts. Reciprocity and exchange rules explain why treatment by an attributed partner may generate felt obligations and reciprocal responses, yet exchange constructs require clear specification of actors, resources, and processes [6]. A psychological contract is narrower: it concerns perceived reciprocal obligations, not simply favorable treatment or a generalized expectation that an organization should be fair. Accordingly, an algorithmic cue enters the proposed contract process only when it is attributed to the organization and interpreted as carrying reciprocal significance.

The theory locates the primary contract at the individual level. Psychological-contract fulfilment perceptions can acquire shared contextual meaning when employees experience common practices, but individual and collective phenomena must remain analytically distinct [7]. Repeated exposure to a common scheduling, evaluation, or allocation system may produce convergent interpretations, yet a team average cannot establish that any particular employee inferred or experienced the same obligation. A future construct such as an algorithmic psychological-contract climate would therefore require evidence of within-unit agreement, common exposure, and multilevel effects rather than simple aggregation.

Algorithms are treated as organizational cue carriers and action enactors within authorized socio-technical arrangements, not as autonomous moral, legal, or relational partners. This preserves the distinction between technological capability and legitimate authority: a system may classify, recommend, constrain, or decide, but the organization remains responsible for granting that role and structuring recourse. The distinction also clarifies voice. Voice opportunity, actual expression, and consequential influence are not interchangeable, while silence can reflect perceived risk, futility, dependence, or channel design [8]. An appeal button may permit expression without enabling review, correction, or accountable justification. Consequential influence is therefore proposed as a stronger procedural condition than nominal participation, although it need not entail employee veto power.

Fairness expectations toward algorithmic systems

Fairness expectations are anticipated standards by which employees evaluate algorithm-mediated outcomes, procedures, explanations, and treatment. They may concern the distribution of opportunities or burdens, the relevance and consistency of decision procedures, the dignity of interaction, or the adequacy of information. Importantly, statistical debiasing and perceived procedural justice can diverge. Reducing persons to standardized inputs may be experienced as procedurally inadequate when relevant individuating information is excluded, even when the stated objective is to reduce bias [9]. Perceived unfairness, however, does not itself demonstrate discriminatory system performance.

Employee reactions also depend on the kind of managerial task delegated to an algorithm. Perceived fairness, trust, and emotion can vary according to decision source and the extent to which a task appears to require contextual, relational, or distinctly human judgment [10]. An algorithm used for routine allocation may therefore be evaluated differently from one used for hiring, discipline, promotion, or dismissal. These task-contingent responses caution against treating "algorithmic fairness" as a stable attitude toward technology or assuming that lower trust automatically signals a perceived broken obligation.

The broader evidence indicates that perceived algorithmic fairness is multidimensional and sensitive to outcomes, procedures, explanations, decision domains, and individual differences [11]. Distributive, procedural, interpersonal, and informational considerations may be salient in different combinations. Outcome favorability, prior trust, technology familiarity, and domain expectations can also shape judgments. Consequently, neither a single fairness score nor a technical fairness metric can substitute for understanding the standards employees actually apply. Perceived fairness must also remain distinct from legal compliance, statistical parity, model accuracy, and organizational legitimacy.

Algorithmic Psychological-Contract Theory therefore treats fairness expectations as possible—but not automatic—obligation content. Established psychological-contract theory requires perceived reciprocal obligation rather than a general evaluative preference [1]. Algorithmic management can supply organizationally attributable exchange cues, but not every cue acquires promissory force [2]. A fairness expectation becomes contract-relevant in the proposed account when organizational claims, prior dealings, or enacted practices reasonably support an inference that the organization owes a particular standard of treatment. Thus, an employee may judge a decision unfair without perceiving breach, or may perceive breach because a specific fairness-related obligation was inferred and then left unfulfilled.

Transparency claims and inferred obligations

Transparency claims include organizational assertions or design cues that an algorithmic system, its decision logic, or its review process is objective, understandable, consistent, accountable, or open to scrutiny. Such claims may matter because algorithm-based human-resource decisions can threaten personal integrity when they reduce contextual understanding, constrain self-presentation or participation, and obscure responsibility [12]. The evidence supports these as ethical and organizational risks, not as universal outcomes. The proposed theory adds that a transparency claim may become a promissory cue when employees attribute it to the organization and interpret it as an assurance about how consequential decisions will be made or reviewed.

Transparency should not be modeled as an isolated quantity of disclosure. Perceptions of fairness, accountability, and transparency can operate jointly in evaluations of algorithmic systems [13]. A technically visible process may still appear illegitimate if responsibility is unclear, data cannot be challenged, or no actor can justify the decision. Conversely, limited disclosure may sometimes coexist with credible accountability and meaningful review. The relevant construct is therefore not maximum visibility, but whether communicated openness is usable within the employment relationship. Cross-sectional associations among these perceptions do not establish that transparency causes fairness, trust, or acceptance.

Explainability adds a further distinction. Explainability and perceived causal intelligibility can influence perceptions, trust, and acceptance, but explanation quality is not equivalent to the mere availability of information [14]. Employees may receive a feature list, score, or generalized rationale without understanding why their case produced a particular outcome. Comprehension also does not guarantee contestability: a person may understand a decision yet lack the authority, evidence access, or safe channel required to challenge it. Acceptance is similarly not evidence of fulfilment, because compliance may reflect dependence, low alternatives, or anticipated futility.

The proposed inference sequence is therefore conditional. A formal claim or interface cue must be attributed to the organization, interpreted as reciprocal and obligation-bearing, and connected to enacted practice. Visibility, comprehensibility, accountability, contestability, and consequential influence should remain separate. Their combination may strengthen the perceived promissory force of transparency, but no element proves that the organization intended or accepted an obligation. This distinction prevents formal policy from being confused with enacted practice and prevents an explanation from being treated as a substitute for correction, human acknowledgment, or structural change.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for definitions, constructs, and conceptual boundaries for algorithmic psychological contracts and fairness expectations in artificial intelligence–mediated work.

Table 1. Definitions, constructs, and conceptual boundaries for Algorithmic Psychological Contracts and Fairness Expectations in Artificial Intelligence–Mediated Work.

Construct or component	Definition	Problem addressed	Distinction from adjacent concepts	Uncertainty	Boundary statement
Organizational algorithmic claim	Formal or enacted signal conveyed through policy, interface, model output, or managerial use of an algorithm and attributed to the organization.	Identifies the cue from which obligation inference may begin.	Not a legal promise or technical function alone.	Promissory force may vary.	Displayed claims do not prove shared obligation.
Algorithmic psychological contract	Employee's perceived reciprocal organizational obligations whose formation, fulfilment, or breach is materially mediated by algorithms.	Extends contract theory without relocating accountability to a machine.	Narrower than general AI attitudes; broader than one fairness judgment.	Central original proposed construct.	Dimensionality and discriminant validity require testing.	The algorithm is not an autonomous contracting party.
Inferred algorithm-mediated obligation	Perceived organizational duty derived from claims and enacted algorithmic practices.	Explains obligation formation without direct human promises.	Requires reciprocity and organizational attribution, unlike preference or hope.	Proposed bridge between cues and contract content.	Employees may infer different duties from identical cues.	Inference does not establish organizational intent or legal commitment.
Fairness expectation	Anticipated distributive, procedural, interpersonal, or informational justice standard.	Specifies evaluation standards in consequential encounters.	Not automatically a psychological-contract obligation.	Possible obligation content when linked to reciprocal claims.	Salience varies by task and context.	Unfairness alone does not establish breach.
Transparency claim	Assertion or design cue that a system or review process is understandable, accountable, or scrutinizable.	Separates communicated openness from usable understanding.	Visibility differs from explanation, control, and influence.	Possible promissory cue and appraisal resource.	Quality, timing, and user capability matter.	Transparency cannot substitute for correction or appeal.
Voice opportunity	Chance to express concerns, provide information, or challenge a decision.	Captures access to expression during formation or repair.	Distinct from whether input changes a decision.	Potential procedural resource.	Risk and futility may suppress use.	Nominal channels do not demonstrate safety or influence.
Consequential influence	Capacity of employee input to trigger review, correction, redesign, or accountable justification.	Prevents symbolic participation from being coded as repair.	More demanding than voice access or interface interactivity.	Proposed condition for credible repair.	Required forms may differ by decision.	Influence does not imply employee veto power.
Perceived algorithmic breach	Appraisal that an organizational obligation materially mediated by an algorithm was not fulfilled.	Links discrepancy recognition to contract theory.	Different from technical error, dissatisfaction, unfairness, or low trust.	Central proposed appraisal state.	Attribution may be divided across actors.	The subtype is proposed and not empirically validated.
Repair adequacy	Perceived sufficiency of procedural, relational, structural, and participatory responses.	Organizes responses beyond explanation alone.	Not equivalent to acceptance, apology satisfaction, or error correction.	Proposed moderator of contract updating.	Different breaches may require different repair combinations.	Effectiveness requires longitudinal comparative evidence.

Note. Evidence bases identify approved sources already introduced in the text. "Original proposed synthesis" marks constructs or relationships advanced by this article rather than established or validated findings.

Perceived algorithmic breach

Perceived algorithmic breach is proposed as an employee appraisal that an organizational obligation, materially communicated or enacted through an algorithmic arrangement, has not been fulfilled. AI adoption has been associated with psychological-contract, engagement, and trust assessments, demonstrating that implementation can enter employees' exchange interpretations [15]. The evidence is cross-sectional and does not show that AI adoption causes breach. Breach requires more than unfavorable technology attitudes: an inferred obligation, a consequential encounter, a discrepancy between anticipated and enacted treatment, and attribution of responsibility to the organization must coincide.

Algorithm implementation is also an organizational meaning-making process. Symbolic actions, local practices, and changing regimes of expertise shape what counts as legitimate knowledge and who appears authorized to act [16]. Thus, the same output may be interpreted as a technical limitation, managerial choice, policy enactment, or organizational repudiation of a prior

commitment. Formal documentation cannot settle that attribution when everyday practice communicates a different allocation of discretion or accountability. The proposed construct therefore concerns experienced organizational non-fulfilment, not an autonomous machine's intention.

Claim–practice discrepancy is especially visible where apparent autonomy coexists with algorithmic surveillance, ratings, competitive discipline, insecurity, or work intensification [17]. Such conditions do not establish breach by themselves; they may instead reflect employment precarity, labor-market structure, or work design. In the proposed theory, discrepancy becomes contract-relevant only when employees had inferred a reciprocal commitment concerning autonomy, fairness, dignity, transparency, or voice. Perceived algorithmic breach must consequently remain distinct from technical error, dissatisfaction, injustice, low trust, and psychological-contract violation, the latter referring to the emotional experience that may follow breach appraisal.

Emotional and behavioral consequences

Psychological-contract breach has longitudinal associations with employee well-being and career-related behavior, and occupational future time perspective can condition these relationships [18]. Applied to algorithm-mediated work, this suggests that identical discrepancies may produce different reactions depending on perceived alternatives, future opportunity, and dependence. The transfer remains proposed because the cited evidence is not algorithm-specific. Immediate disappointment, career reconsideration, and guarded participation should therefore be examined separately rather than combined as a single outcome.

Sustained breach may also relate to longer-term mental and physical health through effort–reward imbalance [19]. This mechanism is relevant when algorithmic systems intensify effort, destabilize reward expectations, or make reciprocity appear persistently unequal. It does not support attributing health harm to one automated decision. Workload, insecurity, organizational change, and prior health can produce similar patterns and must remain rival explanations.

Breach recognition can develop through interconnected triggers involving attention, appraisal, attribution, fairness, coping, and available resources rather than through one isolated event [20]. In algorithm-mediated work, shifting rules, unexplained scores, repeated data errors, or ineffective appeals may become connected into an organizational storyline. Breach reactions and outcomes are temporally unfolding and contingent on resources, interpretation, and interconnected triggers [18-20]. The proposed sequence allows fulfilment, ambiguity, or breach; it does not impose a numerical threshold or presume inevitable escalation.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for proposed mechanisms, relationships, and organizational consequences in algorithmic psychological contracts and fairness expectations in artificial intelligence–mediated work.

Table 2. Proposed mechanisms, relationships, and organizational consequences in Algorithmic Psychological Contracts and Fairness Expectations in Artificial Intelligence–Mediated Work.

Mechanism or relationship	Antecedent	Mediator or process	Moderator	Expected consequence	Rival explanation	Validation requirement
Claim-to-obligation inference	Organizational statements, interfaces, prior exchange history, and enacted treatment	Interpretation of cues as reciprocal duties attributable to the organization	Cue specificity; organizational attribution; prior contract schema	Proposed formation or revision of obligations	General expectation or technology attitude without contract content	Validate measures separating obligation, preference, and fairness expectation.
Fairness-based legitimacy appraisal	Consequential allocation, evaluation, or scheduling	Distributive, procedural, interpersonal, and informational justice appraisal	Decision stakes; task type; comparison source	Proposed fulfilment, ambiguity, or breach appraisal	Outcome favorability, self-interest, or unfamiliarity	Manipulate procedure independently of outcome in field and experimental designs.
Transparency-to-contestability gap	Disclosure, explanation, or visibility claim	Comparison of promised intelligibility with comprehension, appeal, and correction capacity	AI literacy; explanation quality; human review	Proposed credibility gain, ambiguity, or breach	Complexity or information overload	Measure claims, comprehension, and consequential contestability separately.
Trigger accumulation	Repeated discrepancies, shifting rules, or inconsistent treatment	Temporal linkage of triggers into a breach storyline	Severity; prior fulfilment; resource loss	Proposed escalation of breach salience	General change or one adverse event	Use event-based longitudinal research and narrative tracing.
Responsibility attribution	Algorithmic decision or workflow	Allocation of authorship, authority, and accountability across actors	System role; oversight; organizational messaging	Proposed breach target and response choice	Anthropomorphism or limited technical understanding	Vary agency configurations while holding outcomes constant.

Affect and trust recalibration	Consequential perceived non-fulfilment	Emotion, uncertainty, and updating of future-treatment beliefs	Future orientation; dependency; prior trust	Proposed anger, anxiety, guarded compliance, or reduced trust	Job insecurity, workload, or wider change	Combine repeated measures with temporal and rival-path controls.
Voice, silence, resistance, or withdrawal	Perceived breach and available channels	Assessment of safety, efficacy, cost, and responsiveness	Power; status; collective resources	Proposed voice, silence, circumvention, disengagement, or exit	Control intensity or labor-market conditions	Compare employment settings using observed behavior.
Repair and contract updating	Acknowledged discrepancy or challenged decision	Procedural correction, relational acknowledgment, structural change, and consequential influence	Timing; reversibility; repeated enactment	Proposed restoration, narrowing, transformation, or depletion	Acceptance based on convenience or dependence	Compare repair bundles longitudinally with explanation-only responses.

Note. Expected consequences and integrated ordering are original proposed synthesis unless identified as reported evidence. The table presents no effect estimates, validated thresholds, or universal outcomes.

Proposed algorithmic psychological-contract theory

Algorithmic Psychological-Contract Theory proposes a temporal sequence linking organizational cues to exchange updating. Algorithmic management can redesign autonomy, feedback, task variety, demands, and social characteristics, creating a work-design pathway that may operate independently of contract appraisal [21]. Empirical tests must therefore distinguish reactions to changed job characteristics from reactions to perceived non-fulfilment. Work design is an input and rival explanation, not automatic contract content.

Trust is positioned as an antecedent, appraisal resource, consequence, and possible feedback mechanism rather than a proxy for fairness or fulfilment. Trust in AI varies with system characteristics, task conditions, performance, transparency, and embodiment [22]. Prior trust may affect whether a claim is believed, while post-encounter trust can be recalibrated by fulfilment, ambiguity, or breach. Low trust alone cannot establish that an obligation existed.

The proposed theory locates enactment within conjoined human–technology agency. Organizational action can arise from configurations of human discretion, technological constraint, data, and workflow rather than from one isolated actor [23]. Employees may nevertheless attribute responsibility differently across managers, organizations, vendors, and systems. The algorithm is therefore a socio-technical enactor and cue carrier, not an autonomous contracting party with independent legitimate authority.

Algorithmic management may enable and constrain autonomy and value creation in different configurations [24]. The theory accordingly permits positive fulfilment, unresolved ambiguity, and perceived breach rather than assuming uniform harm. Its principal proposed chain is: organizational claims and enacted practices; inferred algorithm-mediated obligations; consequential encounter; discrepancy, attribution, and legitimacy appraisal; fulfilment, ambiguity, or breach; emotional, trust-related, and behavioral response; and contract updating through repeated experience or repair.

Figure 1 presents the original conceptual synthesis for algorithmic psychological-contract theory, showing how the article's principal constructs, mechanisms, uncertainties, and decision boundaries are connected.

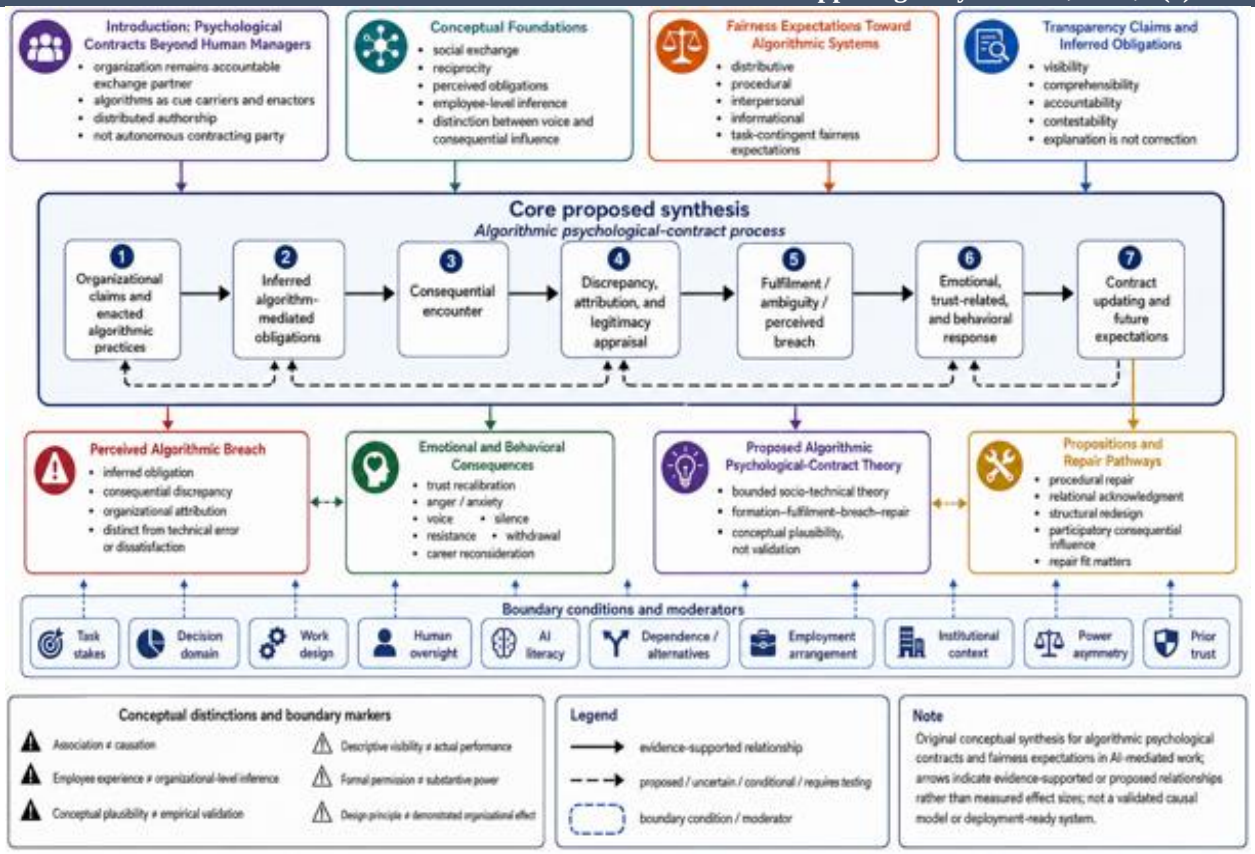

Figure 1. Algorithmic Psychological-Contract Theory

Propositions and repair pathways

The propositions specify relationships requiring testing, not established effects. Limited ability to modify an algorithmic output can reduce aversion in forecasting contexts [25]. Transferred cautiously to employment, **Proposition 1** proposes that usable, consequential opportunities to correct data, supply context, or trigger accountable review will weaken the relationship between discrepancy and perceived breach more than cosmetic control. The proposition requires field tests because modifying a forecast is not equivalent to contesting an employment decision.

Negative reactions are not inevitable. People may prefer algorithmic over human judgment under some conditions [26]. **Proposition 2** therefore proposes that perceived competence, consistency, and task appropriateness will increase the likelihood that algorithm-mediated treatment is interpreted as fulfilment, whereas high relational or contextual demands will strengthen scrutiny of organizational obligations. Algorithm appreciation remains a rival response; preference for advice does not establish legitimate authority, fairness, or contract fulfilment.

Fair employee-management AI requires attention to distributive, procedural, and interactional justice and mechanisms for redressing unfairness [27]. Building on that agenda, **Proposition 3** proposes that repair fit matters: procedural correction should address decision error; relational acknowledgment should address dignity or responsibility failures; structural change should address recurring data, model, or workflow problems; and participatory consequential influence should address denied voice. **Proposition 4** proposes that combined repair will support more favorable contract updating when breach spans multiple obligation domains. These pathways are an original taxonomy, not validated interventions.

Algorithm responses and repair may depend on user agency, judgment source, and justice-sensitive opportunities to modify, contest, or influence the decision process [25-27]. Repair should therefore be assessed through changed practice, accountable explanation, reversibility, timing, and repeated performance rather than acceptance alone. Dependence may produce compliance without restored trust, and an apology may acknowledge harm without correcting the structure that produced it. Testing requires longitudinal field comparisons, qualitative process tracing, experiments that separate outcome from procedure, discriminant measures, and multilevel analysis across occupations and institutional settings.

Figure 2 presents the original conceptual synthesis for mechanisms, boundary conditions, and organizational consequences, showing how the article's principal constructs, mechanisms, uncertainties, and decision boundaries are connected.

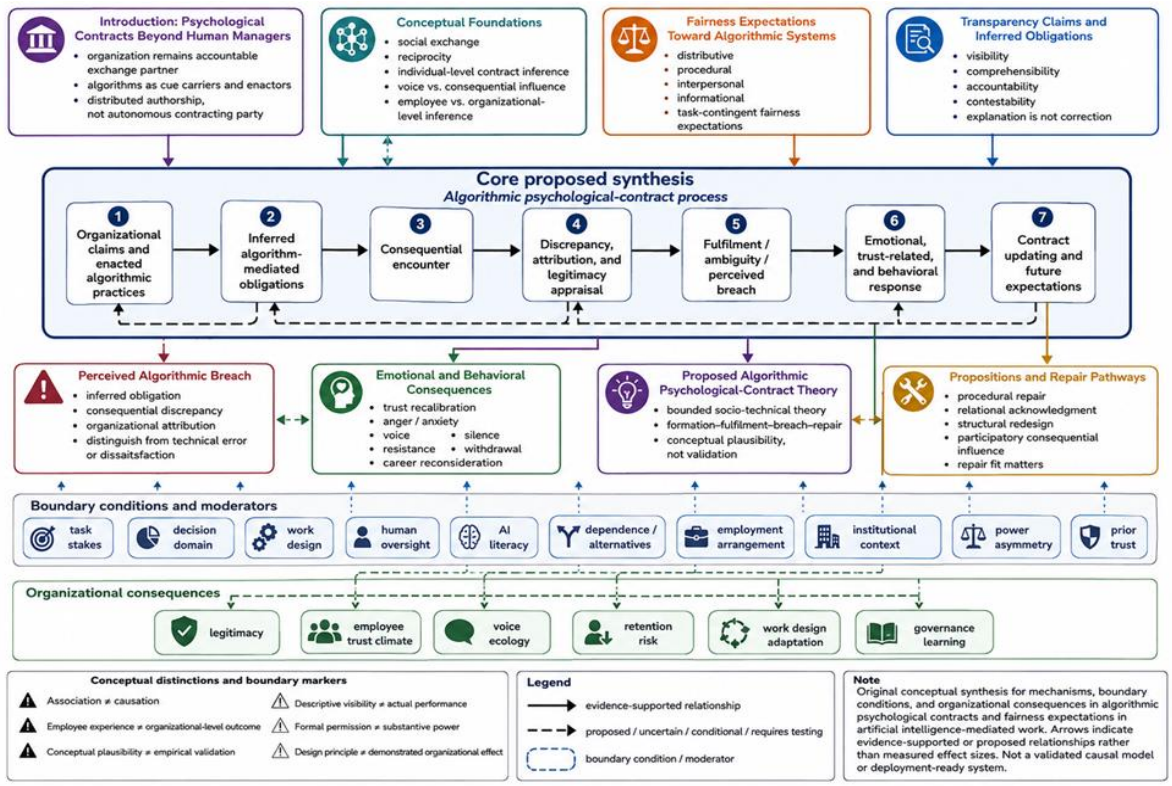

Figure 2. Mechanisms, Boundary Conditions, and Organizational Consequences

Conclusion

Algorithmic Psychological-Contract Theory explains how organizational obligations may be inferred, enacted, fulfilled, or breached when artificial intelligence mediates consequential work decisions. Its central contribution is to preserve the organization as the accountable exchange partner while recognizing that claims, authority, and treatment can be distributed across managers, systems, data, and workflows. Fairness and transparency enter the psychological contract only when employees interpret them as reciprocal obligations; neither perceived injustice nor opacity alone proves breach.

The theory also distinguishes voice access from consequential influence, technological capability from legitimate authority, and acceptance from repair. It proposes that discrepancy, attribution, legitimacy, emotion, and trust updating connect algorithm-mediated encounters to divergent responses, while work design, task stakes, dependence, oversight, literacy, employment arrangements, and institutional context bound those relationships. Procedural, relational, structural, and participatory repair pathways should be evaluated for fit rather than treated as interchangeable solutions.

These constructs, propositions, and pathways remain original conceptual proposals. Their value depends on discriminant measurement and empirical tests capable of separating contract processes from justice, trust, insecurity, work redesign, and general technology attitudes. The theory is neither a validated causal model nor a universal prescription, but a bounded framework for investigating reciprocal expectations beyond exclusively human management.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Coyle-Shapiro JAM, Pereira Costa S, Doden W, Chang C. Psychological contracts: Past, present, and future. Annu Rev Organ Psychol Organ Behav. 2019;6:145-69. doi:10.1146/annurev-orgpsych-012218-015212
2. Tomprou M, Lee MK. Employment relationships in algorithmic management: A psychological contract perspective. Comput Hum Behav. 2022;126:106997. doi:10.1016/j.chb.2021.106997

3. Kellogg KC, Valentine MA, Christin A. Algorithms at work: The new contested terrain of control. Acad Manag Ann. 2020;14(1):366-410. doi:10.5465/annals.2018.0174
4. Raisch S, Krakowski S. Artificial intelligence and management: The automation–augmentation paradox. Acad Manag Rev. 2021;46(1):192-210. doi:10.5465/amr.2018.0072
5. Parker SK, Grote G. Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Appl Psychol. 2022;71(4):1171-204. doi:10.1111/apps.12241
6. Cropanzano R, Anthony EL, Daniels SR, Hall AV. Social exchange theory: A critical review with theoretical remedies. Acad Manag Ann. 2017;11(1):479-516. doi:10.5465/annals.2015.0099
7. Tekleab AG, Laulié L, De Vos A, De Jong JP, Coyle-Shapiro JAM. Contextualizing psychological contracts research: A multi-sample study of shared individual psychological contract fulfilment. Eur J Work Organ Psychol. 2020;29(2):279-93. doi:10.1080/1359432X.2019.1608294
8. Morrison EW. Employee voice and silence: Taking stock a decade later. Annu Rev Organ Psychol Organ Behav. 2023;10:79-107. doi:10.1146/annurev-orgpsych-120920-054654
9. Newman DT, Fast NJ, Harmon DJ. When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. Organ Behav Hum Decis Process. 2020;160:149-67. doi:10.1016/j.obhdp.2020.03.008
10. Lee MK. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data Soc. 2018;5(1):2053951718756684. doi:10.1177/2053951718756684
11. Starke C, Baleis J, Keller B, Marcinkowski F. Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. Big Data Soc. 2022;9(2):20539517221115189. doi:10.1177/20539517221115189
12. Leicht-Deobald U, Busch T, Schank C, Weibel A, Schafheitle S, Wildhaber I, et al. The challenges of algorithm-based HR decision-making for personal integrity. J Bus Ethics. 2019;160(2):377-92. doi:10.1007/s10551-019-04204-w
13. Shin D, Park YJ. Role of fairness, accountability, and transparency in algorithmic affordance. Comput Hum Behav. 2019;98:277-84. doi:10.1016/j.chb.2019.04.019
14. Shin D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. Int J Hum Comput Stud. 2021;146:102551. doi:10.1016/j.ijhcs.2020.102551
15. Braganza A, Chen W, Canhoto A, Sap S. Productive employment and decent work: The impact of AI adoption on psychological contracts, job engagement and employee trust. J Bus Res. 2021;131:485-94. doi:10.1016/j.jbusres.2020.08.018
16. Pachidi S, Berends H, Faraj S, Huysman M. Make way for the algorithms: Symbolic actions and change in a regime of knowing. Organ Sci. 2021;32(1):18-41. doi:10.1287/orsc.2020.1377
17. Wood AJ, Graham M, Lehdonvirta V, Hjorth I. Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. Work Employ Soc. 2019;33(1):56-75. doi:10.1177/0950017018785616
18. Zacher H, Rudolph CW. Relationships between psychological contract breach and employee well-being and career-related behavior: The role of occupational future time perspective. J Organ Behav. 2021;42(1):84-99. doi:10.1002/job.2495
19. Griep Y, Bankins S, Vander Elst T, De Witte H. How psychological contract breach affects long-term mental and physical health: The longitudinal role of effort–reward imbalance. Appl Psychol Health Well Being. 2021;13(2):263-81. doi:10.1111/aphw.12246
20. Wiechers HE, Coyle-Shapiro JAM, Lub XD, ten Have S. The tremors of interconnected triggers over time: How psychological contract breach can erupt. J Organ Behav. 2022;43(7):1172-89. doi:10.1002/job.2645
21. Parent-Rocheleau X, Parker SK. Algorithms as work designers: How algorithmic management influences the design of jobs. Hum Resour Manag Rev. 2022;32(3):100838. doi:10.1016/j.hrmr.2021.100838
22. Glikson E, Woolley AW. Human trust in artificial intelligence: Review of empirical research. Acad Manag Ann. 2020;14(2):627-60. doi:10.5465/annals.2018.0057
23. Murray A, Rhymer J, Sirmon DG. Humans and technology: Forms of conjoined agency in organizations. Acad Manag Rev. 2021;46(3):552-71. doi:10.5465/amr.2019.0186
24. Meijerink J, Bondarouk T. The duality of algorithmic management: Toward a research agenda on HRM algorithms, autonomy and value creation. Hum Resour Manag Rev. 2023;33(1):100876. doi:10.1016/j.hrmr.2021.100876
25. Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Manage Sci. 2018;64(3):1155-70. doi:10.1287/mnsc.2016.2643
26. Logg JM, Minson JA, Moore DA. Algorithm appreciation: People prefer algorithmic to human judgment. Organ Behav Hum Decis Process. 2019;151:90-103. doi:10.1016/j.obhdp.2018.12.005
27. Robert LP Jr, Pierce C, Marquis L, Kim S, Alahmad R. Designing fair AI for managing employees in organizations: A review, critique, and design agenda. Hum Comput Interact. 2020;35(5-6):545-75. doi:10.1080/07370024.2020.1735391