E-ISSN: 3108-4192

APSSHS

Academic Publications of Social Sciences and Humanities Studies
2024, Volume 4, Issue 2, Page No: 93-102
Available online at: https://apsshs.com/

Journal of Applied Organizational Systems and Behavior

Artificial Intelligence and the Reproduction of Organizational Inequality

Charlotte Dupuis[1]*, Hugo Perrin[1], Elise Morel[2], Thomas Martin[3]

1. Department of AI and Inequality, Faculty of Business, University of Lille, Lille, France.
2. Department of Organizational Inequality Reproduction, Faculty of Management, University of Montpellier, Montpellier, France.
3. Department of Social Justice at Work, Faculty of Psychology, University of Nantes, Nantes, France.

Abstract

Artificial intelligence is increasingly used to allocate work, evaluate performance, screen applicants, distribute opportunities, and structure managerial attention. Yet organizational debate often isolates technical bias from implementation, worker participation, error distribution, and accountability, obscuring how disadvantage can persist across an entire decision process. This narrative review examines how artificial intelligence may reproduce organizational inequality through four connected domains: bias in data and system design, participation deficits during implementation, unequal exposure to automated error, and accountability failure. A structured, evidence-bounded review approach combined explicit questions, relevance-focused searches, transparent screening, critical reading, thematic comparison, and PRISMA-compatible reporting. The literature converges on the importance of data provenance, proxy and objective choice, authority allocation, information asymmetry, contestability, and remedy capacity. It conflicts, however, over whether observed disparities arise primarily from encoded social history, optimization processes, institutional conditions, user interpretation, or pre-existing organizational inequality. Evidence is also heterogeneous across computational audits, field studies, experiments, qualitative process research, conceptual analyses, and review articles. The article therefore proposes an original multistage synthesis in which design choices shape implementation participation; participation shapes exposure and challenge capacity; accountability conditions correction; and organizational responses may feed into later data and evaluation. This synthesis is not a validated causal model. Its implications are conditional on occupation, employment relation, institutional context, worker resources, organizational governance capacity, and time. Longitudinal, multilevel, comparative, and intervention-based research is required to test, refine, or reject the proposed relationships.

Keywords: Inequality in artificial intelligence–mediated organizations, Participation deficits during implementation, Unequal exposure to automated error, Accountability failure, Reproduction; Inequality

How to cite this article: Dupuis C, Perrin H, Morel E, Martin T. Artificial Intelligence and the Reproduction of Organizational Inequality. J Appl Organ Syst Behav. 2024;4(2):93-102. https://doi.org/10.51847/3o8nZO48aI

Received: 06 February 2024; **Revised:** 15 March 2024; **Accepted:** 17 March 2024
Corresponding author: Charlotte Dupuis
E-mail ✉ charlotte.dupuis@univ-lille.fr

Introduction

Artificial intelligence increasingly determines who is noticed, selected, scheduled, rewarded, monitored, developed, or removed. Its inequality implications cannot be reduced to whether AI is fairer than human judgment. AI can redistribute workplace opportunity and control unevenly across individual, team, organizational, and institutional levels, with effects that vary by task, technology, and worker position [1, 2]. Inequality in AI-mediated organizations is therefore defined here as patterned asymmetry in access, visibility, evaluation, authority, exposure to error, and capacity for correction. It includes distributive outcomes and the processes through which some actors become more legible, influential, or contestable.

The prevailing technical-bias framing is necessary but incomplete. Algorithmic systems can direct tasks, evaluate conduct, discipline workers, and structure the information available to managers, while affected actors may accommodate, contest, or

strategically respond to those controls [3]. Similar tools can produce different implications depending on employment status, dependence, system knowledge, managerial discretion, and collective representation. Interface neutrality does not settle whose data and knowledge became variables, who may challenge an inference, or who bears error.

Nor can automation and augmentation be treated as mutually exclusive organizational futures. They are interdependent choices about the distribution of tasks, discretion, expertise, and authority between humans and machines [4]. An ostensibly augmentative system may still narrow employee judgment if recommendations become compulsory in practice, whereas an automated process may retain meaningful review when reviewers possess information, time, jurisdiction, and correction authority. Technological capability differs from legitimate authority, and nominal human involvement differs from consequential oversight.

Learning systems further complicate static accounts because their outputs can reshape organizational knowledge, coordination, and subsequent practice [5]. Data are not merely historical inputs; they may also be products of earlier classifications, worker adaptations, managerial priorities, and unequal access to opportunities. When these outputs influence future assignments or evaluations, inequality may be reproduced through recursive organizational processes rather than a single biased decision. The direction and strength of this recursion remain empirical questions.

This narrative review develops an original multistage explanation linking design, implementation participation, unequal exposure to error, accountability, and feedback. Its contribution is to distinguish these domains while showing why they must be analyzed together: workplace inequality under AI is heterogeneous and may be reduced, displaced, reproduced, or intensified depending on how capabilities and control are distributed [1]. The review separates descriptive evidence from causal inference, worker perceptions from organizational outcomes, voice opportunity from influence, and formal policy from enacted practice. The proposed framework organizes evidence and testable questions; it does not establish a universal causal sequence or an implementation-ready governance system.

Narrative review method

The article used a structured narrative-review approach aligned with its theory-organizing purpose. Review questions addressed how design bias, implementation participation, error exposure, accountability, and recursive organizational processes are defined, connected, contested, and bounded. Searches combined title variants with terms concerning workplace inequality, algorithmic management, recruiting, data and proxy bias, worker voice, error distribution, explainability, accountability, governance, multilevel conditions, measurement, and validation. Explicit search, selection, critical-reading, and thematic-synthesis procedures distinguish a structured review from an informal literature summary [6].

Figure 1 reports the verified flow of records through identification, duplicate removal, screening, full-text assessment, exclusion, and final inclusion for this narrative review article.

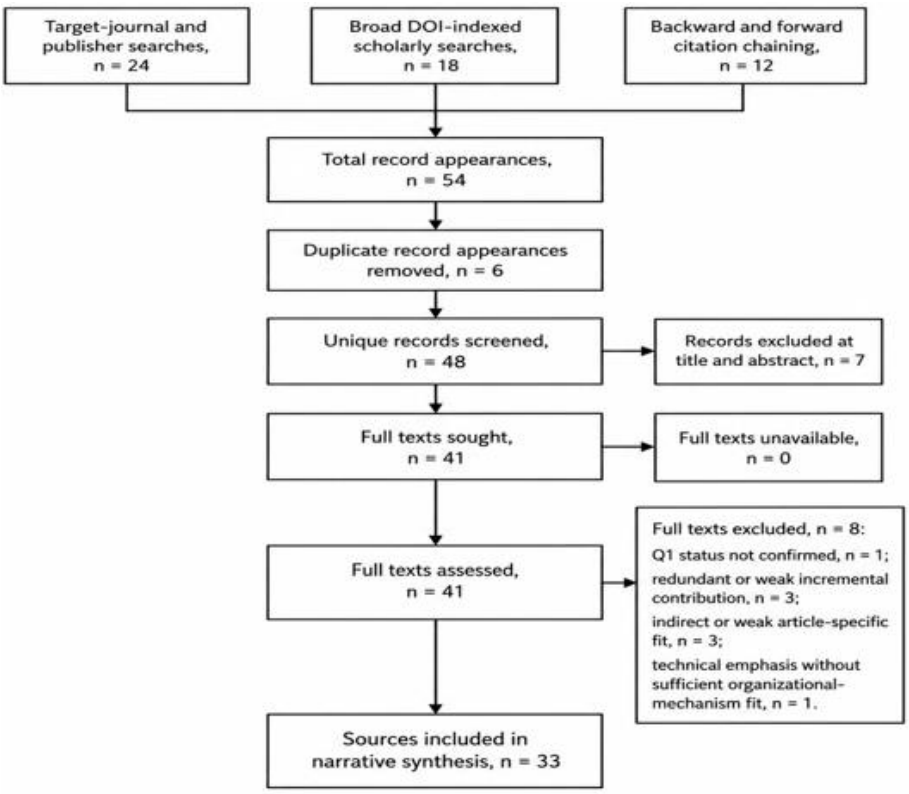

Figure 1. PRISMA Flow Diagram for Evidence Identification and Selection

Eligible sources were peer-reviewed journal articles with verified metadata, DOI information, relevant journal standing, and direct support for a planned construct, mechanism, boundary, table element, figure component, or research priority. Review quality depends on coherent problem formulation, search logic, conceptual organization, and contribution alignment rather than mechanical accumulation of sources [7]. Duplicate DOI or repeated article appearances were removed before screening. Title-and-abstract review excluded records without sufficient article-specific relevance; full-text assessment considered design, context, evidence contribution, transferability, and risk of overclaiming.

PRISMA-compatible reporting was used to document the selection ledger, not to redefine the article as a systematic effectiveness review [8]. Extraction recorded study purpose, unit, setting, data source, time horizon, construct or measure, reported relationship, alternative explanation, limitation, transferability boundary, and exact claim permitted. No common effect metric or quality score was imposed across heterogeneous designs.

Synthesis proceeded through critical comparison of convergence, conflict, mechanisms, levels, contextual boundaries, and unresolved uncertainty. Review articles may advance theory by clarifying constructs, assumptions, and relationships, but a review-derived framework remains distinct from empirically validated theory [9]. Accordingly, the selected literature is purposefully bounded [6]. The method supports transparent, evidence-bounded interpretation; it does not establish exhaustive capture, pooled effects, or causal validation.

Bias in data and system design

Bias in data and system design refers here to value-laden choices about data provenance, labels, proxies, objectives, categories, representations, thresholds, and delivery processes that can shape organizational visibility or allocation. It is broader than a coding defect and narrower than every unequal outcome associated with technology. Algorithmic HR design can implicate personal integrity while learned representations can encode historical social associations [10, 11]. These findings establish that normative choices and social histories may enter technical systems, but encoded association does not itself prove an adverse employment decision, discriminatory intent, or a single downstream causal pathway.

Table 1 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for review design, evidence domains, and extraction framework for artificial intelligence and the reproduction of organizational inequality.

Table 1. Review design, evidence domains, and extraction framework for Artificial Intelligence and the Reproduction of Organizational Inequality.

Evidence domain or review construct	Review question	Eligible evidence type	Extraction fields	Appraisal or relevance criterion	Synthesis role	Main limitation	Interpretation boundary
Structured narrative review	How was evidence bounded?	Method scholarship [6].	Query; source; claim; limitation	Direct methodological fit	Makes selection explicit	Nonexhaustive	PRISMA reports flow only [8].
Data and system design	Which choices shape visibility?	Computational, field, ethical, review evidence	Data; label; proxy; objective; group	Identifiable design process	Separates encoding from consequence	Heterogeneous outcomes	Association does not prove discrimination [10].
Implementation participation	Who influences implementation?	Qualitative and work-system evidence	Actor; expertise; stage; authority	Consequential influence	Examines knowledge and power	Original review construct	Consultation is not equitable implementation.
Automated error and accountability	Who receives and remedies error?	Audits, experiments, governance analyses	Error; subgroup; explanation; appeal	Distribution or remedy mechanism	Connects accuracy, legitimacy, recourse	Original review construct	Transparency does not establish correction.
Inequality reproduction	How might disadvantage recur?	Longitudinal, comparative, critical evidence	Allocation; adaptation; feedback; time	Process or boundary mechanism	Connects decisions across time	Proposed integration [9].	The cycle requires testing.
Corrective governance	Which control matches each risk?	Decision, audit, participation research	Actor; authority; trigger; correction	Control mapped to risk	Organizes stage-matched principles	Proposed architecture	Principles do not guarantee outcomes.

Note. Later-domain rows are original extraction constructs. Their substantive evidence is introduced only in the corresponding sections, preserving citation order.

Optimization can also produce unequal exposure without an explicit protected-class rule. In a field study of STEM career advertising, a nominally gender-neutral campaign was delivered less often to women because the platform's optimization operated within gender-differentiated advertising costs [12]. The result demonstrates an organizationally relevant allocation pathway between campaign design and labor-market visibility. It does not isolate deliberate discrimination, and competing explanations include pricing, platform objectives, audience behavior, and market competition. Consequently, audit of explicit variables alone may miss disparities generated by delivery systems.

AI-enabled recruiting scholarship identifies recurring concerns involving discrimination, construct validity, opacity, privacy, applicant dignity, and the legitimacy of organizational decision processes [13]. These categories converge on the need to evaluate the full recruiting pathway rather than a model in isolation. Yet the evidence base combines conceptual argument, technical evaluation, perceptions, and field applications, so it does not establish equal prevalence or magnitude across technologies and occupations. Human judgment is also not a bias-free comparator; pre-existing recruitment practices may shape both training data and the organizational baseline against which AI is assessed.

The proposed synthesis therefore treats design as an allocation of organizational visibility and value. Choices about what is recorded, predicted, optimized, and ignored can structure later control and contestation [3]. This connection is analytically plausible but not yet a measured causal chain. Design bias may be attenuated, amplified, or redirected during implementation, especially when domain experts, workers, managers, vendors, and affected applicants differ in information access and authority.

Participation deficits during implementation

Participation deficits are exclusions from consequential influence over problem definition, data interpretation, workflow redesign, testing, authority allocation, revision, and withdrawal. Participation is not attendance at meetings or permission to provide feedback. Ethnographic evidence from AI development for hiring shows that technical and recruitment experts negotiated how professional knowledge would be translated into machine-recognizable categories [14]. Such translation can shape what the system treats as valid expertise. The case is situated, not universal, and participation cannot guarantee equity.

Implementation also changes which forms of knowledge appear legitimate. Longitudinal qualitative evidence shows that symbolic accommodation to an algorithm can coexist with preserved prior practices and still produce unanticipated changes in an organization's regime of knowing [15]. Machine-readable indicators may acquire authority because they appear consistent, scalable, or objective, while situated judgment becomes harder to defend. Workers differ in their ability to translate experience into accepted data, but qualitative evidence does not estimate this mechanism's prevalence or causal magnitude.

Participation deficits are especially visible in platform and app-mediated work. Algorithmic management can centralize task allocation, monitoring, evaluation, and sanction while weakening conventional employment-relations channels through which workers question decisions [16]. Nominal flexibility or formal contractor status does not establish practical autonomy when access to work, ratings, and income depends on opaque platform rules. These implications depend on occupation, platform dependence, employment classification, labor institutions, and collective representation; they do not transfer automatically to conventional organizations.

More generally, implementation participation is a work-system property involving information access, intelligibility, contestability, voice channels, and allocation of decision rights. Information asymmetry and technical-organizational opacity shape how workers experience and respond to algorithmic management [17]. Formal permission to speak differs from power to alter a model, workflow, or decision. Consequential participation requires that relevant actors can access evidence, articulate situated knowledge, receive a reasoned response, and influence revision through recognized authority.

Participation can nevertheless be unequal even when an organization invites consultation. Technical expertise, professional jurisdiction, vendor dependence, managerial sponsorship, time, collective resources, and employment security affect whose knowledge enters implementation decisions [14]. The evidence does not establish that one participation mechanism prevents inequality. Research must distinguish voice opportunity from consequential influence and assess whose concerns alter system objectives, testing criteria, authority arrangements, and correction practices over time.

Unequal exposure to automated error

Unequal exposure to automated error occurs when groups or organizational positions differ in the likelihood, consequence, detection, or correction of mistaken classification. Aggregate accuracy can conceal this distribution. In population-health management, using expenditure as a proxy for health need understated the needs of Black patients because spending reflected unequal access and treatment as well as illness burden [18]. The study demonstrates a specific proxy mismatch, not the prevalence of the same mechanism across employment systems.

Explanation and human review are therefore conditional resources rather than automatic safeguards. Evidence from an AI-supported candidate-management system suggests that explanations and recommendations can alter selection behavior, but their contribution depends on what information is revealed and how reviewers interpret it [19]. An explanation can rationalize

a flawed proxy, overload a reviewer, or remain inconsequential when the reviewer lacks time or correction authority. Explainability should not be equated with individuation, accuracy, or remedy.

Error also has a procedural and relational dimension. Experimental evidence indicates that fairness, trust, and emotion vary across algorithmic-management tasks because people make different assumptions about where machines or humans possess appropriate competence [20]. These responses matter for acceptance, cooperation, and willingness to challenge a decision, but perceived unfairness is not proof of discriminatory allocation. Conversely, favorable perceptions cannot establish subgroup accuracy or equitable consequences.

Applicants may judge algorithm-only hiring as less fair when they believe it cannot recognize personal uniqueness, although judgments vary with procedure and outcome favorability [21]. Such findings identify legitimacy conditions, not long-term labor-market effects. The review therefore defines unequal exposure as the interaction of error distribution with consequence, detection resources, and challenge capacity. Proxy-driven allocation provides direct evidence for one component [18]; the complete construct remains an original synthesis requiring field-based and longitudinal testing.

Table 2 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for cross-study findings, mechanisms, convergence, and conflict in artificial intelligence and the reproduction of organizational inequality.

Table 2. Cross-study findings, mechanisms, convergence, and conflict in Artificial Intelligence and the Reproduction of Organizational Inequality.

Construct, mechanism, or relationship	Representative context	Reported finding	Evidence design	Convergent evidence	Conflicting evidence	Methodological concern	Review interpretation
Historical association	Learned language representations [11].	Social associations entered embeddings [11].	Computational association tests	Recruiting scholarship identifies data-bias risks [13].	Association does not establish an employment outcome [11].	Translation from representation to decision is indirect.	Upstream risk signal, not proof of consequence.
Optimization and exposure	STEM career-ad delivery [12].	Delivery differed without explicit gender targeting [12].	Field observational study	Proxy choice also produced unequal allocation [18].	Market and audience processes remain alternatives [12].	One causal pathway was not isolated.	Examine objectives and delivery, not code alone.
Expert translation	AI development for hiring [14].	Technical and professional knowledge were negotiated [14].	Qualitative ethnography	Implementation altered organizational knowing [15].	Jurisdiction and local change may also explain outcomes.	Context-specific process evidence.	Assess influence over assumptions, not meeting attendance.
Proxy mismatch	Health-management allocation [18].	Cost understated need for one racial group [18].	Retrospective audit	Optimization shaped differential labor-market visibility [12].	Domains and mechanisms differ.	Prevalence across workplace AI is unknown.	Audit subgroup error, consequence, and correction access.
Explanation and legitimacy	Candidate and management decisions [19].	Explanation changed decision processes [19].	Prototype evaluation	Fairness and trust varied by task [20].	Applicant responses varied by procedure [21].	Perception may not predict field outcomes.	Explanation requires accuracy, authority, and remedy.

Note. Review interpretations are original synthesis; no pooled effect or cross-domain equivalence is claimed.

Accountability failure

Accountability failure is a break between system visibility and the ability to obtain an effective organizational response. Transparency may expose documentation or outputs while leaving causal relations, institutional responsibilities, and opportunities for intervention unclear [22]. A worker can see a score yet remain unable to identify the relevant data, responsible actor, review standard, or correction route. Opacity is thus one failure mode, not the complete accountability problem.

Responsibility is also displaced when organizational actors describe algorithmic output as autonomous or inevitable. Ethical analysis places responsibility on organizations and designers for the roles delegated to algorithms and for decision structures that prevent individuals from exercising responsibility [23]. This does not settle legal liability. It instead rejects the inference that technical delegation dissolves organizational authorship, especially where objectives, procurement choices, workflow rules, and response procedures remain humanly arranged.

Accountable AI requires institutions able to demand reasons, interrogate evidence, challenge outcomes, and impose correction [24]. A nominal human in the loop may provide little protection when that person lacks expertise, jurisdiction, independence, or permission to depart from the recommendation. The appropriate arrangement differs across public agencies, employers, and platforms, but the analytical test is similar: who must answer, to whom, on what evidence, and with what consequence?

People analytics illustrates why accountability must extend beyond model explanation. The literature identifies surveillance, privacy intrusion, reductionism, power asymmetry, and misinterpretation as recurring organizational and employee risks [25]. Their prevalence and magnitude are not uniform, and analytics may also support legitimate organizational purposes. Accountability nevertheless fails when affected people bear monitoring or classification consequences while responsibility is fragmented across managers, vendors, analysts, and policies.

The proposed accountability chain links visibility, intelligibility, answerability, review authority, accessible remedy, correction, and feedback closure [22]. Each link is distinct: disclosure may not yield understanding; understanding may not create authority; and appeal availability may not produce timely correction. This chain is an original diagnostic synthesis rather than a validated scale. Empirical evaluation should examine actual use, independence, timeliness, reversal, recurrence, and unequal access to each stage.

Mechanisms of organizational inequality reproduction

Organizational inequality may be reproduced when evaluation changes behavior in ways that become inputs to later evaluation. Longitudinal process evidence shows that workers facing opaque algorithmic assessments infer hidden rules and alter their conduct around those inferences, creating an "invisible cage" of individualized adaptation [26]. Workers with greater time, information, financial security, or peer support may adapt differently, but the study does not estimate a complete self-reinforcing causal loop.

Nominal autonomy can coexist with intense control. Comparative evidence from global platform work shows that temporal and spatial flexibility may accompany insecurity, algorithmic direction, and dependence on ratings or work allocation [27]. Under such conditions, responsibility for remaining visible and available shifts toward workers, while the organization retains important informational and evaluative power. Transferability depends on occupation, alternatives, employment institutions, and income dependence.

Classification also shapes how organizational differences become interpreted. Data infrastructures rank people and convert prior behavior and social position into market-like signals that can appear deserved or objective [28]. Applied to AI-mediated work, this perspective suggests that scores may not merely describe performance; they can influence access, status, and subsequent opportunities. This is a plausible structural mechanism, not direct evidence that every AI score produces cumulative disadvantage.

Fairness metrics remain valuable for detecting specified disparities, but their scope is limited. A narrow antidiscrimination discourse may leave category construction, institutional power, and the production of disadvantage outside the analysis [29]. Metric parity should therefore be combined with inquiry into whose attributes become comparable, which outcomes matter, and how burdens accumulate. This boundary does not justify abandoning quantitative audit; it limits what parity alone can establish.

Table 3 organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for evidence limitations, unresolved questions, and future research priorities.

Table 3. Evidence limitations, unresolved questions, and future research priorities.

Context or level	Moderator or boundary condition	Evidence pattern	Organizational implication	Alternative explanation	Transferability limit	Uncertainty	Research need
Applicant or employee	Task, procedure, outcome favorability [20].	Fairness and trust vary by context [21].	Pair perception with outcome audit.	Novelty or prior beliefs.	Experiments may not predict careers.	Perception–distribution relation unresolved.	Link field outcomes, procedures, and subgroup responses.
Platform employment	Dependence, alternatives, institutions [27].	Autonomy can coexist with control [27].	Judge voice against actual dependence.	Labor demand or income need.	Platform occupations differ from standard employment.	Adaptive costs are weakly quantified.	Comparative panels and natural experiments.
Implementation team	Expertise, jurisdiction, decision rights [14].	Knowledge translation shapes encoded expertise [14].	Measure consequential influence by decision stage.	Wider organizational change [15].	Qualitative settings are context dependent.	Common participation measures are absent.	Process tracing and configuration comparison.

Decision architecture	Stakes, uncertainty, expertise [4].	AI redistributes authority and coordination [5].	Match decision rights to risk.	Complementary resources may drive outcomes.	Conceptual structures need field tests.	Effective oversight configurations remain unsettled.	Field experiments and decision-log analysis.
Data and proxy	Provenance, objective, subgroup base rates [18].	Distinct representation and proxy risks [11].	Separate construct, allocation, error, and remedy audits.	History, market optimization, or measurement error [12].	Technologies are not homogeneous.	Drift and intersectional error remain sparse.	Construct validation, subgroup audits, and appeal linkage.
Accountability institution	Mandate, access, independence [24].	Disclosure may remain disconnected from remedy [22].	Specify actor, authority, trigger, and closure.	Incentives may dominate formal principles.	Sectors differ in disclosure and enforcement.	Remedy use and impact remain incomplete.	Compare regimes and track correction and recurrence.
Temporal structure	Adaptation, retraining, turnover [26].	Classification may shape later data and opportunity [28].	Monitor feedback across decision cycles.	External market or policy change.	Most evidence observes few stages.	Lag and reversal conditions are uncertain.	Linked longitudinal and multilevel designs.

Note. Organizational implications and research needs are original proposed synthesis, not validated prescriptions.

The reproduction mechanism proposed here connects allocation, worker adaptation, generated data, organizational interpretation, and later evaluation [26]. It is conditional rather than inevitable: contestability, mobility, collective resources, managerial discretion, and correction may interrupt the cycle. Pre-existing inequality may also cause both system inputs and observed outcomes, while labor-market or policy changes may mimic feedback. The framework must therefore be tested against alternative temporal and multilevel explanations rather than inferred from disparity alone.

Corrective design and governance principles

Corrective governance should begin with decision architecture rather than a generic demand for human oversight. Organizations can delegate decisions, sequence human and AI judgments, or aggregate them, with different implications for expertise, contestability, and responsibility [30]. Authority should be matched to stakes, uncertainty, error consequences, and review capacity. Human presence is not meaningful oversight unless the person can access relevant evidence, exercise independent judgment, and initiate correction.

Figure 2 presents the original conceptual synthesis for integrated evidence synthesis framework for artificial intelligence and the reproduction of organizational inequality, showing how the article's principal constructs, mechanisms, uncertainties, and decision boundaries are connected.

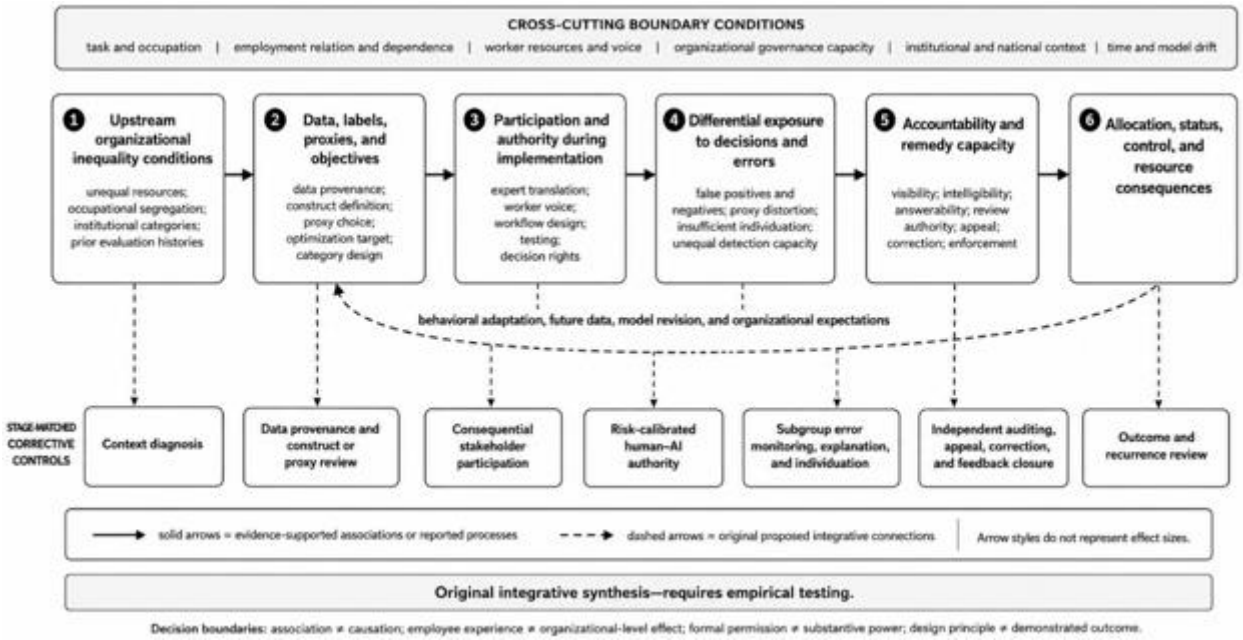

Figure 2. Integrated Evidence Synthesis Framework for Artificial Intelligence and the Reproduction of Organizational Inequality

Auditing, operational tools, and ethics principles can organize AI governance, but their effectiveness depends on scope, access, implementation, and institutional follow-through [31-33]. Audits can be narrow, dependent, or evidentially constrained; tools may not fit organizational authority structures; and principle convergence does not establish enacted practice. Governance evaluation should therefore examine who commissions scrutiny, what evidence is accessible, whether findings trigger correction, and whether affected groups can use the process.

Upstream controls should document data-generating processes and test whether a proxy represents the construct relevant to the organizational decision [18]. Integrity-sensitive design should also identify where categorization removes context or restricts an affected person's capacity to respond [10]. These controls cannot eliminate historical inequality, measurement uncertainty, or strategic behavior. They provide traceability for contestable choices and enable later monitoring of subgroup errors and model drift.

Implementation controls require consequential participation in assumptions, testing, workflow, and decision rights, not consultation detached from authority [14]. Remedy controls require answerability and an institution capable of scrutinizing and changing outcomes [24]. Appeals should be accessible, timely, documented, and linked to system revision when errors recur. Their existence alone does not demonstrate equitable use or distributive improvement.

Testing the framework requires designs capable of separating levels, mechanisms, and time. Workplace effects vary across technologies and positions [1]. Multilevel research must connect individual experience with team, organizational, and labor-market processes without treating them as interchangeable [2]. Longitudinal process evidence is needed to examine adaptation and feedback [26]. Comparative field studies, interventions, audits, linked decision records, and qualitative process tracing should test interruption, displacement, recovery, and unintended effects.

Conclusion

This narrative review argues that artificial intelligence may reproduce organizational inequality through a connected but noninevitable sequence. Data, labels, proxies, and objectives shape what becomes visible; implementation arrangements determine whose knowledge and authority matter; errors are distributed through organizational positions and resources; and accountability determines whether disadvantage can be recognized and corrected. Worker adaptation and organizational responses may then enter later data and evaluation, creating the possibility of recursive reproduction.

The principal contribution is an original multistage framework that separates technical bias from participation, exposure, accountability, and temporal feedback while showing why isolated assessment is insufficient. Corrective design is therefore framed as stage-matched governance: context diagnosis, construct and proxy review, consequential participation, risk-calibrated decision rights, subgroup monitoring, independent challenge, accessible remedy, and recurrence review.

These principles are analytical proposals rather than validated organizational effects. The reviewed evidence is heterogeneous, context dependent, and often limited to one level or stage. It cannot establish one universal causal pathway, a complete ranking of interventions, or implementation readiness. Longitudinal, multilevel, comparative, and intervention-based studies should test whether the proposed connections occur, for whom, under which institutional conditions, and whether corrective arrangements reduce, relocate, or inadvertently intensify inequality.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Karunakaran A, Lebovitz S, Narayanan D, Rahman H. Artificial intelligence at work: An integrative perspective on the impact of AI on workplace inequality. Acad Manage Ann. 2025;19(2):693-735. doi:10.5465/annals.2023.0230
2. Bankins S, Ocampo AC, Marrone M, Restubog SLD, Woo SE. A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. J Organ Behav. 2024;45(2):159-82. doi:10.1002/job.2735
3. Kellogg KC, Valentine MA, Christin A. Algorithms at work: The new contested terrain of control. Acad Manage 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 Manage Rev. 2021;46(1):192-210. doi:10.5465/amr.2018.0072

5. Faraj S, Pachidi S, Sayegh K. Working and organizing in the age of the learning algorithm. Inf Organ. 2018;28(1):62-70. doi:10.1016/j.infoandorg.2018.02.005
6. Snyder H. Literature review as a research methodology: An overview and guidelines. J Bus Res. 2019;104:333-9. doi:10.1016/j.jbusres.2019.07.039
7. Paul J, Criado AR. The art of writing literature review: What do we know and what do we need to know? Int Bus Rev. 2020;29(4):101717. doi:10.1016/j.ibusrev.2020.101717
8. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372. doi:10.1136/bmj.n71
9. Post C, Sarala RM, Gatrell C, Prescott JE. Advancing theory with review articles. J Manag Stud. 2020;57(2):351-76. doi:10.1111/joms.12549
10. 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
11. Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science. 2017;356(6334):183-6. doi:10.1126/science.aal4230
12. Lambrecht A, Tucker C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Manage Sci. 2019;65(7):2966-81. doi:10.1287/mnsc.2018.3093
13. Hunkenschroer AL, Luetge C. Ethics of AI-enabled recruiting and selection: A review and research agenda. J Bus Ethics. 2022;178(4):977-1007. doi:10.1007/s10551-022-05049-6
14. van den Broek E, Sergeeva A, Huysman M. When the machine meets the expert: An ethnography of developing AI for hiring. MIS Q. 2021;45(3):1557-80. doi:10.25300/MISQ/2021/16559
15. 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
16. Duggan J, Sherman U, Carbery R, McDonnell A. Algorithmic management and app-work in the gig economy: A research agenda for employment relations and HRM. Hum Resour Manag J. 2020;30(1):114-32. doi:10.1111/1748-8583.12258
17. Jarrahi MH, Newlands G, Lee MK, Wolf CT, Kinder E, Sutherland W. Algorithmic management in a work context. Big Data Soc. 2021;8(2):20539517211020332. doi:10.1177/20539517211020332
18. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-53. doi:10.1126/science.aax2342
19. Hofeditz L, Clausen S, Rieß A, Mirbabaie M, Stieglitz S. Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. Electron Mark. 2022;32(4):2207-33. doi:10.1007/s12525-022-00600-9
20. 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
21. Lavanchy M, Reichert P, Narayanan J, Savani K. Applicants' fairness perceptions of algorithm-driven hiring procedures. J Bus Ethics. 2023;188(1):125-50. doi:10.1007/s10551-022-05320-w
22. Ananny M, Crawford K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc. 2018;20(3):973-89. doi:10.1177/1461444816676645
23. Martin K. Ethical implications and accountability of algorithms. J Bus Ethics. 2019;160(4):835-50. doi:10.1007/s10551-018-3921-3
24. Busuioc M. Accountable artificial intelligence: Holding algorithms to account. Public Adm Rev. 2021;81(5):825-36. doi:10.1111/puar.13293
25. Giermindl LM, Strich F, Christ O, Leicht-Deobald U, Redzepi A. The dark sides of people analytics: Reviewing the perils for organisations and employees. Eur J Inf Syst. 2022;31(3):410-35. doi:10.1080/0960085X.2021.1927213
26. Rahman HA. The invisible cage: Workers' reactivity to opaque algorithmic evaluations. Adm Sci Q. 2021;66(4):945-88. doi:10.1177/00018392211010118
27. 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
28. Fourcade M, Healy K. Seeing like a market. Socioecon Rev. 2017;15(1):9-29. doi:10.1093/ser/mww033
29. Hoffmann AL. Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. Inf Commun Soc. 2019;22(7):900-15. doi:10.1080/1369118X.2019.1573912
30. Shrestha YR, Ben-Menahem SM, von Krogh G. Organizational decision-making structures in the age of artificial intelligence. Calif Manage Rev. 2019;61(4):66-83. doi:10.1177/0008125619862257
31. Mökander J, Morley J, Taddeo M, Floridi L. Ethics-based auditing of automated decision-making systems: Nature, scope, and limitations. Sci Eng Ethics. 2021;27(4):44. doi:10.1007/s11948-021-00319-4
32. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics. 2020;26(4):2141-68. doi:10.1007/s11948-019-00165-5

33. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99. doi:10.1038/s42256-019-0088-2