

International Theory and Practice in Humanities and Social Sciences

doi: 10.70693/itphss.v3i6.745
Vol 3, No 6 (2026)
ISSN 3079-4579 eISSN 3078-4387

OPEN ACCESS ARTICLE

How Organizational AI Ethics Governance Shapes Employee Creativity: A Conceptual Analysis of Mechanisms and Boundary Conditions

Bing Zhang[1] and Linpei Song[1,*]

[1]Hunan Institute of Science and Technology, 414006, China
*Corresponding author: Linpei Song 12023036@hnist.edu.cn

Abstract

As AI becomes embedded in organizational decision-making and knowledge production, employees gain new tools for accessing and generating information while also confronting concerns about algorithmic evaluation, data use, and accountability. Drawing on research on AI ethics, algorithmic management, and organizational creativity, this conceptual study defines organizational AI ethics governance as the rules, procedures, and allocation of responsibilities that guide the development, deployment, operation, and retirement of AI systems. We argue that such governance can shape employee creativity through several connected mechanisms. Fair and contestable procedures can strengthen procedural justice, while transparent system boundaries and meaningful human review can support calibrated trust. Tolerance for responsible experimentation, together with discretion over how AI is used, can protect psychological safety and job autonomy. Governance also determines whether human–AI collaboration broadens exploration or channels employees toward standardized outputs. These effects depend on task characteristics, the stakes of AI-assisted decisions, whether AI is used primarily for automation or augmentation, employees' AI literacy, and the consistency between formal rules and managerial conduct. This study suggests that organizations should translate ethical principles into risk-based review, purpose limitation in data use, effective contestability and appeal mechanisms, employee participation, and oversight across the AI system life cycle so that AI can serve as a resource for careful creative work rather than a substitute for judgment.

Key words: organizational AI ethics governance; employee creativity; procedural justice; calibrated trust; psychological safety; human–AI collaboration

Introduction

The digitalization of organizations is changing how information informs decisions and how employee performance is evaluated. AI systems are now closely integrated into recruitment screening, performance appraisal, information retrieval, and content generation. The Ethical Norms for the New Generation Artificial Intelligence call for fairness, impartiality, privacy, security, reliability, and accountability throughout the governance, development, production, and use of AI (National Governance Committee for the New Generation Artificial Intelligence, 2021). Applying these principles within organizations raises practical questions. What information does a model use to reach a judgment? Can affected employees contest that judgment? May data collected for testing be repurposed as evidence of performance, and how much discretion do employees retain? Any analysis of AI and creativity must therefore consider the organizational rules that govern the technology.

Organizational research commonly defines creative ideas as both novel and useful, and frontline suggestions and improvements often provide the starting point for organizational innovation (Anderson et al., 2014). AI expands the range of information that employees can search and can generate numerous candidate outputs within a short period. At the same time, it can entrench performance metrics and encourage employees to prioritize measurable targets over questioning established assumptions. Creative work is especially difficult to evaluate in advance and often entails trial and error, dissent, and temporary failure. When technology increases employees' sense of being monitored or their fear of retrospective blame, gains in efficiency may not translate into greater willingness to engage in creative work.

Existing research offers several perspectives on these issues, but the relevant conversations remain insufficiently integrated. AI ethics research tends to emphasize principles, risks, and social acceptance; research on algorithmic management focuses mainly on control and employees' experience of work; and creativity research has largely examined managerial support, job design, and team climate. Fairness, explainability, privacy, and human oversight are often discussed separately rather than as elements of organizational governance. Generative AI introduces a further tension: employees can produce initial solutions quickly, yet those solutions may converge on similar forms and patterns of thought. Better individual output does not necessarily yield stronger organizational innovation. This article therefore examines how organizational AI ethics governance shapes employees' creative engagement and the conditions under which its effects may differ.

© The Author 2026(BY CC 4.0). International Theory and Practice in Humanities and Social Sciences (ITPHSS) journal homepage: wisvora.org/index.php/itphss

Linking Organizational AI Ethics Governance to Employee Creativity

Organizational AI ethics governance is more than a statement of values. It comprises the rules, procedures, and allocation of responsibilities that apply throughout an AI system's development, deployment, operation, and retirement. Although AI ethics frameworks originate from different institutions, transparency, fairness, accountability, privacy, and harm prevention recur across them (Jobin et al., 2019). Within organizations, these principles must be translated into concrete practices, including scrutiny of data sources, assessment of disparate effects across groups, specification of a model's appropriate scope, human review of consequential decisions, and clear responsibility for system failures. Governance principles can influence behavior only when employees can observe how these arrangements operate in their everyday work.

From an employee perspective, effective governance includes several interdependent elements. Procedures should guard against unjustified discrimination and allow affected employees to provide information that a system has missed. Explanations should address decisions that affect a person's role or rights rather than merely offering general technical guidance. Human oversight should give reviewers the authority to pause a process or change an outcome. Data governance should also specify the purpose of collection, permitted uses, and retention periods. Responsibility among business managers, technical staff, vendors, and users should remain distinct, while affected employees should be able to participate in developing the rules. An organization may increase pressure rather than reduce it if it offers explanations without accepting challenges, or nominally provides human review while denying reviewers the authority to intervene.

Creative work differs from routine tasks because its outcomes are difficult to specify in advance and its evaluation criteria vary across time and contexts. Prior research indicates that a supportive work environment facilitates creative performance (Amabile et al., 1996). Introducing AI adds activities such as query formulation, prompt design, output selection, and fact-checking to the workflow. Employees may use models to extend the boundaries of their knowledge, but they may also reproduce familiar model outputs repeatedly. Whether AI-generated material counts as an original contribution, who bears responsibility for errors, and how the material's professional value should be assessed often remain opaque. The way organizational rules address these questions can determine whether employees use AI to explore possibilities or confine themselves to routine tasks that are easy to measure.

Mechanisms Connecting AI Ethics Governance and Employee Creativity

Procedural Justice and Calibrated Trust

Employees judge a procedure as fair based on more than whether the same algorithm is applied to everyone. They also consider whether the information is accurate, whether the procedure can be corrected, and whether they have an opportunity to express their views. In human resource decisions, limited transparency about AI can weaken employees' perceptions of procedural justice (Pei et al., 2021). Algorithms may reduce some forms of managerial bias, yet a system that reduces a person's full experience to a numerical score can itself compromise procedural justice (Newman et al., 2020). This problem bears directly on creativity. Because undeveloped ideas are easy to dismiss, employees who do not understand the evaluation criteria, or who fear that a single mistake will remain permanently attached to them, are likely to favor predictable work.

An effective explanation should help employees understand and, where necessary, correct a decision. Organizations need not disclose every technical detail of a model, but they should identify the data sources used, the main job-relevant criteria, plausible sources of error, and available avenues for appeal. Decisions concerning recruitment, promotion, and performance appraisal directly affect employees' interests. Affected employees should therefore be able to provide information that a model cannot recognize and have their objections reviewed by a person with decision authority. These safeguards cannot eliminate every problem, but they can make adverse outcomes open to discussion. For creative work, procedures that allow employees to voice concerns and verify decisions reduce the perceived cost of departing from established practice. By contrast, formally consistent procedures that cannot correct errors may encourage employees to adapt their behavior to what they believe the AI system rewards.

Trust provides a second mechanism. Research suggests that the reliability of AI outputs, system transparency, and the purpose for which AI is used all influence human trust (Glikson and Woolley, 2020). The goal of governance, however, is not to maximize AI use. It is to help employees judge when reliance is appropriate for the task and its risks. Low-risk tasks with readily verifiable outputs may permit greater reliance on automation. When information is incomplete, values conflict, or substantial novelty is required, employees need to retain a questioning stance. Organizations can help employees distinguish useful from unreliable suggestions by communicating the scope of training data, relevant model changes, and common failure modes. Conversely, managers create distrust and defensive behavior when they require employees to follow an automated process but assign all responsibility for errors to the user.

Calibrated trust is an adjustable relationship in which employees may follow an AI recommendation, reject it, request further information, or suspend use. Managers should set the frequency and depth of review according to risk instead of treating a system output as a reason to avoid responsibility. Review, on this account, does not express blanket opposition to AI; it is a form of deliberate judgment. Employees can use time saved by AI to reconsider problems, compare alternatives, and verify important facts, thereby treating AI as a knowledge aid rather than a substitute for thought.

Psychological Safety and Job Autonomy

Creative engagement requires employees to acknowledge uncertainty and voice ideas that are not yet fully developed. Research suggests that team members are more willing to experiment and exchange ideas when interpersonal risk does not expose them to ridicule or punishment (Edmondson, 1999). Under algorithmic management, however, automated scoring, continuous records, and the linkage of data across contexts expand what organizations can observe and may intensify workloads or restrict discretion (Parent-Rocheleau and Parker, 2022). If the process of exploration becomes part of a punitive performance system, employees may hide mistakes and repeat familiar practices, reducing their willingness to undertake risky experiments. Employees can confront failure honestly only when organizations distinguish the reporting of problems from the assignment of blame; under those conditions, psychological safety can support continued exploration.

Job autonomy concerns employees' discretion over task sequence, work methods, and professional judgments. Self-determination theory suggests that satisfying the need for autonomy helps sustain work motivation (Gagné and Deci, 2005). AI can relieve employees of repetitive work and offer alternatives, but mandatory settings and real-time feedback can also prescribe a single way of working. Governance should align control with accountability. Employees cannot exercise meaningful autonomy when they remain nominally responsible for outcomes but lack the authority or information needed to judge them.

Human–AI Collaboration and the Dual Effects of AI

Human–AI collaboration rests on differences in capability. AI is well suited to rapid search, statistical pattern detection, and the generation of initial options. Humans remain responsible for interpreting context, distinguishing value conflicts, and deciding whether a recommendation should be adopted. Automation and augmentation are not independent paths; both usually appear in everyday organizational activities (Raisch and Krakowski, 2021). Experimental research indicates that generative AI can improve the speed and quality of certain writing tasks (Noy and Zhang, 2023). Other evidence suggests that AI assistance may raise creativity ratings for individual outputs

while making outputs produced by different people more similar (Doshi and Hauser, 2024). Working faster at the individual level, therefore, does not necessarily strengthen an organization's capacity for innovation.

When performance systems emphasize efficiency, employees have stronger incentives to accept the high-probability suggestions supplied by AI. Output may rise in the short term, but less common perspectives, knowledge, and skills may receive less attention. AI plays a different role when an organization values problem definition, evidence gathering, the comparison of alternatives, and fit with the specific context. Under those conditions, AI can supply material that employees challenge and rework rather than simply directing the course of innovation. Asking employees to explain why they accepted one recommendation and rejected others can develop their judgment and help the organization detect emerging patterns of innovation. Human–AI collaboration should therefore vary across stages of the workflow. A collaborative design neither rejects AI nor delegates final judgment to it. Its purpose is to help employees complete their work while continuing to develop the ability to redefine problems and evaluate alternatives.

Boundary Conditions and Governance Implications

The effects of governance depend mainly on task characteristics and how AI is used. Employees may be more willing to accept AI for tasks with clear standards and readily identifiable errors. Decisions about career opportunities, the evaluation of innovative work, and resource allocation have higher stakes and require contextual judgment; they therefore call for stronger explanation, review, and correction. For automation-oriented AI, governance must address control, accountability allocation, and error correction. For augmentation-oriented AI, it should clarify the system's appropriate scope, require scrutiny of information, and recognize human contributions. In content production, employees may adapt their work to an algorithm if an organization treats output volume or model-generated scores as the sole performance standard. When a model instead provides options and employees retain final decision authority, they have more room to exercise creativity.

Employees' AI literacy also conditions the effects of governance. Failure to understand a model's limitations may produce resistance or uncritical reliance. Training should therefore cover not only technical skills but also bias detection, factchecking, data responsibilities, copyright rules, and assessment of appropriate use. Employees also interpret formal policies in light of an organization's past conduct. In a company dominated by intensive performance monitoring and one-way directives, even a transparent policy may be understood as a new form of surveillance. Individual governance measures are unlikely to earn employees' trust when formal rules conflict with managerial behavior.

Organizations can translate these considerations into four areas of practice. First, they can classify systems according to the scope and severity of potential effects, requiring bias testing and workforce impact assessment before deploying high-risk systems. Second, they can provide concise explanations for different forms of algorithmic decisions, together with contestability and appeal mechanisms and access to human review. Third, they can separate learning records from performance records and define access rights, retention periods, and prohibited purposes for creative files, prompts, and internal knowledge. Fourth, they can involve employees from different roles in selection and feedback, assessing whether AI improves the quality of creative work rather than measuring processing speed alone. These practices can help employees evaluate AI-assisted work more accurately.

Governance should cover the full AI system life cycle. At project initiation, organizations should define the intended purpose and identify affected groups. During procurement or development, they should review data provenance, test results, and vendor responsibilities. Before deployment, a limited pilot can compare how employees in different roles experience the system. After deployment, organizations should document errors, complaints, and patterns of inappropriate reliance, then revise permissions and controls as needed. System retirement also requires attention to data deletion, data transfer, and the handover of responsibilities. For large AI systems, a joint review group comprising business, technical, human resource, and compliance personnel can reduce situations in which no one explains technical failures, and no one accepts responsibility for management decisions. Employee participation does not transfer responsibility to employees; it uses frontline experience to improve governance.

Conclusion and Future Research

Whether AI supports employee creativity depends on how organizations adopt and govern it. Clear rules, meaningful opportunities to contest decisions, and preserved human judgment can strengthen procedural justice and support calibrated trust. Tolerance for responsible failure can protect psychological safety and job autonomy. A deliberate division of work between people and AI across stages of innovation may also limit the convergence of thought that can accompany rapid production. In contrast, opaque algorithms, excessive surveillance, and unclear accountability may prompt defensive behavior or uncritical dependence on AI. Organizations should therefore translate ethical principles into operational arrangements for risk classification, approval, data governance, and participation. Future research could distinguish formal governance structures from employees' perceptions of those structures, develop organization-level measures of AI ethics governance, and use multilevel designs to examine how individual creativity relates to organizational innovation performance.

Funding

This study was supported by the Scientific Research Project of the Hunan Provincial Department of Education (No. 25B0579).

References

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., and Herron, M. (1996). Assessing the work environment for creativity. *Academy of Management Journal*, 39(5):1154–1184.

Anderson, N., Potočnik, K., and Zhou, J. (2014). Innovation and creativity in organizations: A state-of-the-science review, prospective commentary, and guiding framework. *Journal of Management*, 40(5):1297–1333.

Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):Article eadn5290.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2):350–383.

Gagné, M. and Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, 26(4):331–362.

Glikson, E. and Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2):627–660.

Jobin, A., Ienca, M., and Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399.

National Governance Committee for the New Generation Artificial Intelligence (2021). Ethical norms for the new generation artificial intelligence. [In Chinese], September 25, 2021.

Newman, D. T., Fast, N. J., and Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes*, 160:149–167.

Noy, S. and Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192.

Parent-Rocheleau, X. and Parker, S. K. (2022). Algorithms as work designers: How algorithmic management influences the design of jobs. *Human Resource Management Review*, 32(3):Article 100838.

Pei, J., Liu, S., Zhong, C., and Chen, Y. (2021). Can AI algorithmic decision-making improve employees' perceptions of procedural justice? *Foreign Economics & Management*, 43(11):41–55. [In Chinese].

Raisch, S. and Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1):192–210.