



Generative Artificial Intelligence Governance and Algorithmic Trust in Smart Organizations: Implications for Sustainable Innovation and Organizational Performance

Masoumeh Cheraghi¹, Masoud Fallahi^{2*}, Mohammad Mafi³

¹M.A. in Psychology and B.A. in Business Administration, University of Applied Science and Technology, Ilam Branch, Ilam, Iran

²B.Sc. in Applied Chemistry, Islamic Azad University, South Tehran Branch, Tehran, Iran, CEO of Ravan Gostaran Tavan Gharb Company

³Ph.D. student in Industrial Management (Systems Management), Islamic Azad University, Sanandaj Branch, Sanandaj, Iran; Deputy CEO of Ravan Gostaran Tavan Gharb Company.

Article info

Received: 23.06.2026

Accepted: 15.10.2026

Available Online: 17.10.2026

Checked for Plagiarism: Yes

Keywords:

generative artificial intelligence;
AI governance; algorithmic trust;
smart organizations; responsible
innovation; sustainable
innovation

ABSTRACT

Purpose Generative artificial intelligence (GenAI) is being integrated into knowledge work, organizational decision support, and sustainability-related operations, yet evidence regarding governance, algorithmic trust, and durable organizational value remains fragmented. This article develops an integrated explanation of how responsible governance can enable, rather than merely constrain, the creation of sustainable innovation and organizational performance. Design/methodology/approach A critical, structured integrative synthesis draws on recent peer-reviewed research and authoritative policy and standards publications issued between 2023 and October 2026. Sources compared through thematic coding across governance architecture, calibrated algorithmic trust, human-AI collaboration, sustainable innovation, and organizational performance. The article proposes constructs, testable propositions, an implementation framework, and a future empirical research design; it does not claim to report original survey or experimental findings. Findings Governance quality theorized to improve organizational outcomes through calibrated trust and responsible human-AI practices. The effects are contingent on decision criticality, employee AI literacy, accountability arrangements, institutional pressure, and the environmental costs of computation. Trust appropriately calibrated to system reliability rather than maximized. Practical implications Organizations should combine documented use-case inventories, risk-tiered oversight, evaluation, traceability, worker participation, and multidimensional scorecards. Originality/value the framework connects governance mechanisms and behavioral trust calibration to an explicitly triple-bottom-line account of innovation and performance, while separating established empirical evidence from propositions requiring testing.

Introduction

Generative artificial intelligence (GenAI) has shifted from a peripheral experiment to a general-purpose organizational interface for retrieving,

transforming, generating and evaluating information.

Smart organizations entities that coordinate connected data, adaptive digital systems and human

*Corresponding Author: **Masoud Fallahi** (Masoud.fallahi0918@gmail.com)

1 Email: mcheraghi42@gmail.com

3 Email: Msina222@gmail.com

expertise deploy GenAI in customer engagement, programming, design, human resource management, procurement, product development and environmental reporting.

Unlike conventional predictive analytics, foundation-model applications often act through open-ended dialogue, probabilistic generation, retrieval augmentation and increasingly agentic workflows. Their outputs are plausible even when inaccurate; they also reshape the allocation of cognitive labor and managerial authority. These properties place governance and trust at the center of organization design (Janssen,2025; Luengo Vera et al.,2026).

The contemporary adoption context reinforces the relevance of the problem. The Stanford AI Index 2026 reports widespread organizational AI use, while the 2025 McKinsey survey identifies workflow redesign and governance leadership as important organizational adaptations. Yet diffusion should not be conflated with value realization: reported adoption rates depend on survey definitions, samples and time windows; they do not directly establish causal performance effects (Stanford HAI,2026; McKinsey,2025). Rigorous evidence from a customer-support setting nonetheless demonstrates that GenAI assistance can improve measured task productivity, while its benefits vary considerably across workers and activities (Brynjolfsson et al.,2025).

The core dilemma is that strategic value and material harm may emerge from the same affordances. GenAI can accelerate experimentation, reduce search costs and extend expertise, but it may also produce hallucinations, amplify bias, expose confidential data, narrow the diversity of ideas and impose significant resource burdens. Studies of AI-assisted creativity show improved individual outputs accompanied, in some settings, by reduced collective novelty or diversity (Doshi & Hauser,2024; Meincke et al.,2025). Thus, efficiency is a necessary but insufficient criterion for responsible implementation, particularly when employees and stakeholders are affected differently. Governance scholarship has emphasized frameworks, principles and organizational accountability, while behavioral research has explored trust in algorithmic decisions. Sustainability research, meanwhile, increasingly examines environmental information processing and GenAI-related performance. These streams remain insufficiently connected. For example, a firm might meet procedural disclosure requirements but undermine employees' perceptions of legitimacy; alternatively, a highly trusted chatbot may be used uncritically and degrade decision quality. The governance challenge is not to induce unqualified trust but to construct justified reliance, contestability and the capability to detect and correct errors (Batoool et al.,2025; Schilke & Reimann,2025).

This article addresses three questions.

- ✓ RQ1: Which configurations of GenAI governance are most likely to foster calibrated algorithmic trust in smart organizations?
- ✓ RQ2: Through what socio-technical mechanisms might governance and trust contribute to sustainable innovation and organizational performance?
- ✓ RQ3: Which boundary conditions and trade-offs can reverse or attenuate those relationships? In response, the article develops a multilevel model, propositions, a measurement architecture and a feasible empirical testing agenda.

The contribution is threefold. Theoretically, it integrates risk-based AI governance, socio-technical systems, organizational information processing and trust calibration. Analytically, it distinguishes adoption, responsible use, innovation capability and realized economic, social and environmental performance. Practically, it operationalizes a governance-trust-value pathway with concrete implementation controls and indicators. The article is positioned as a conceptual synthesis rather than as a systematic review with a reproducible exhaustive-search claim or an empirical study with invented data.

Conceptual Foundations and Boundaries

GenAI governance is defined here as the allocation of decision rights, structures, rules, technical safeguards, relational arrangements and assurance activities that guide GenAI acquisition, configuration, use, monitoring and retirement. It includes board accountability, model and data inventories, third-party management, application-specific risk assessment, documented human oversight, red-teaming, provenance controls and incident response. Governance must therefore be treated as an observable organizational capability rather than a statement of ethical aspirations (Mäntymäki and colleagues' governance traditions are extended by Batoool et al.,2025; Janssen,2025).

Algorithmic trust is an actor's willingness to rely on an algorithmic system under uncertainty on the basis of perceptions and evidence of competence, predictability, integrity, benevolence or procedural legitimacy. Organizationally, trust exists at several levels: the individual employee trusts output quality, managers trust controllability, customers trust the fairness of decisions and regulators evaluate demonstrable accountability. These levels are related but should not be collapsed into a single score. Calibrated trust denotes an alignment between actual system capability and the strength or scope of reliance; both under-reliance and over-reliance can be costly.

A smart organization is conceptualized as a socio-technical network integrating data infrastructure,

human expertise, organizational routines, connected services and adaptive analytics. Smartness does not imply that decisions are fully automated or that more automation is inherently superior. It denotes the ability to sense conditions, interpret signals, act adaptively and learn from performance and harm. GenAI can extend interpretive capacity, but human actors retain responsibility for objectives, institutional commitments and exceptional cases. Sustainable innovation involves novel or substantially improved products, services, business models and processes that advance long-run economic viability while accounting for environmental stewardship and social consequences. It is not synonymous with the number of AI-supported ideas produced. Novelty must be

evaluated alongside adoption, diffusion, lifecycle impact, stakeholder welfare and the rebound effects of increased digital activity. Organizational performance similarly comprises operational quality and productivity, financial resilience, workforce outcomes, compliance robustness and sustainability impacts, rather than short-term cost savings alone. The proposed constructs are distinct. Governance provides the control environment, trust influences reliance behavior, responsible use represents enacted work practices, innovation captures change capabilities and performance captures realized outcomes. Distinguishing them enables mediational explanations and guards against circular survey instruments in which respondents are asked essentially the same question under different labels.

Table 1. Construct boundaries and operational definitions

Construct	Core meaning	Distinguishing evidence
Governance maturity	Institutionalized structures, practices and safeguards	Audit trail, assigned owners, risk assessments
Algorithmic trust calibration	Reliance aligned with demonstrated model quality	Acceptance and override on seeded-error tasks
Responsible AI use	Enacted verification, escalation and authorization	Behavioral logs and sampled decisions
Sustainable innovation capability	Generation and implementation of net-beneficial novelty	Validated solutions and lifecycle impact
Organizational performance	Economic, social and environmental realized value	Quality-adjusted ROI and independent indicators

Review Method and Evidence Logic

The research design is a critical integrative review combined with theory development. Its goal is conceptual integration and the identification of testable mechanisms, not estimation of pooled treatment effects or systematic enumeration of all publications. The evidence window prioritizes publications from 2024 through October 2026, supplemented by the foundational AI management system standard released in December 2023. No pre-2023 works are included in the reference list. This inclusion strategy responds to the rapid development of GenAI-specific organizational literature. Literature was identified through targeted searches of scholarly publisher pages and institutional repositories, using combinations of “generative AI”, “governance”, “algorithmic trust”, “organization”, “sustainability”, “innovation”, “performance” and “human–AI collaboration”. Records were retained when their abstracts, methods or institutional status directly informed at least one core construct. Priority was given to identifiable peer-reviewed articles with digital object identifiers and primary publications by public standards or research institutions. Practitioner surveys were used only for contextual claims about deployment and work practices. An interpretive matrix was constructed around five dimensions: theoretical contribution; governance mechanism; trust or human behavior mechanism;

innovation/performance outcome; and limits of the evidence. Contradictory findings were not averaged away. For example, experimental studies suggesting creative gains were read alongside evidence of reduced idea diversity; field productivity benefits were separated from organizational net value; and rules-based governance prescriptions were compared with observed worker circumvention of perceived illegitimate controls. Because the search process was focused rather than fully reproducible and database-exhaustive, no PRISMA flow diagram, formal quality-weighted meta-analysis or total-record screening count is claimed. This methodological boundary is substantively important: it avoids presenting the article as a systematic review without the corresponding registered protocol, duplicate elimination, full search strings and independent screening. The synthesis is intended to generate a coherent research model and a transparent empirical agenda rather than prevalence estimates. The inference hierarchy distinguishes three claim types: (a) observed relationships or effects in cited empirical settings; (b) design guidance from standards and policy documents; and (c) novel propositions advanced in this paper. A source supporting category (a) cannot by itself prove the causal pathway represented in the conceptual figure. Equally, alignment with NIST or ISO guidance

cannot establish effectiveness without implementation evidence. This separation is maintained throughout the findings and propositions.

Table 2. Evidence-synthesis logic and limits

Evidence class	Role in synthesis	Interpretive limitation
Controlled or field empirical studies	Suggest observed effects and heterogeneity	Setting-specific; limited external validity
Peer-reviewed conceptual and review papers	Define constructs and rival mechanisms	No direct causal estimate
Standards / regulatory instruments	Identify guidance and obligations	Conformance does not prove effectiveness
Institutional reports and surveys	Establish context and trends	Often self-reported and noncausal

Governance Architecture for Organizational GenAI

Effective governance should be risk-proportionate, lifecycle-oriented and embedded in workflows. The ISO/IEC 42001:2023 management-system model emphasizes continual improvement, responsibilities and documented processes; the NIST Generative AI Profile organizes context-sensitive risk management within govern, map, measure and manage activities. The updated OECD AI Principles emphasize trustworthy, human-centered development and the ability to address emerging risks (ISO,2023; Autoio et al.,2024; OECD,2024). These instruments are complementary but do not constitute interchangeable certifications or identical legal obligations.

At the structural layer, organizations require executive ownership, a cross-functional governance committee, defined risk appetite and designated accountability for each deployed use case. Procurement decisions must consider model licensing, data location, contractual restrictions, audit rights, change notices and supplier continuity. A high-impact system that drafts employment recommendations warrants markedly different treatment from a low-impact assistant for internal brainstorming. Governance effectiveness arises from prioritization, not uniform bureaucracy.

At the procedural layer, each use case should undergo problem framing, data suitability analysis, impact assessment, approval, testing, deployment monitoring and periodic reauthorization. Evaluation should test not merely average answer quality but distributional failure: vulnerable users, minority-language tasks, long contexts, adversarial prompts, confidential content and stale retrieved material. Escalation pathways must specify when employees must stop reliance, seek expert review or report a

near miss. Evidence of evaluation must be auditable and attributable.

At the relational layer, participation and legitimacy matter. Human oversight that exists only on paper may become a rubber stamp when workload incentives reward speed. Employees should have a credible channel to challenge outputs, describe unsafe pressures and propose improved safeguards without retaliation. Leon (2026) reports organizational evidence that layered guardrails can improve multiple assurance outcomes while adding task-time costs and generating circumvention when perceived as disproportionate. This makes the legitimacy of controls a performance-relevant design feature rather than a peripheral attitude.

Governance must also address dynamic model behavior. External providers update systems, retrieval corpora change, prompts evolve and workarounds spread informally. Therefore, assurance cannot stop at procurement certification or a one-time red-team exercise. Change logs, reproducible evaluation sets, version-aware monitoring, override rates and observed incident patterns should inform governance recalibration. Governance is best understood as an adaptive control loop rather than as a static ethics document. Finally, governance choices should cover environmental externalities. Model selection affects energy demands, latency, infrastructure intensity and vendor concentration. Carbon-aware architecture can prioritize smaller task-specific models, retrieval, caching, batching, responsible data retention and demand reduction when quality requirements permit. These practices must be evaluated using task-normalized impact accounting, not assumed to be sustainable because they are digital. IEA (2025) explicitly addresses both the energy demand of AI and potential efficiency gains across energy systems.

Table 3. Governance controls by lifecycle stage

Stage	Minimum control	Enhanced control for consequential uses
Intake	Use-case inventory and owner	Formal impact assessment and authorization
Procurement	Supplier and data assessment	Audit rights, change clauses, contingency plan
Design	Approved workflow and boundaries	Human override, source grounding, protected data
Predeployment	Benchmark and security review	Disaggregated tests, red-team and adversarial cases
Operations	Feedback and incident log	Continuous monitoring, sampling, escalation
Change / retirement	Version records and closure	Revalidation, preservation and exit testing

Algorithmic Trust as a Calibrated Organizational Capability

Trust is consequential because employees cannot fully audit the training data, latent representations and update behavior of a large language model before every use. Reliance decisions therefore combine prior system experience, interface cues, documented safeguards, peer norms and managerial expectations. Trust is productive when it enables appropriate delegation and efficient verification; it becomes hazardous when confidence exceeds demonstrated capability. A mature organization should optimize calibration and informed contestability rather than indiscriminately attempting to maximize acceptance. Transparency alone does not guarantee trust. Explaining that an output was AI-generated, showing a confidence score or exposing a reasoning narrative may help users in some settings but can mislead when indicators lack empirical validity. Schilke and Reimann (2025) found experimentally that disclosure of AI use may erode trust in the actor who uses it. Sullivan and Weger (2025) examined how different transparency levels affect trust and perceived reliability in AI-supported decisions. These findings imply that transparency should answer actionable questions what the system did, what evidence it used, when it fails and whom to contact rather than merely satisfy a disclosure ritual. Work roles affect calibration. Novices may overvalue fluent responses, while experienced professionals may reject useful support or detect domain-specific errors more readily. AI literacy, substantive expertise and verification tools jointly shape reliance. Brynjolfsson et al. (2025) document substantial heterogeneity in productivity gains, illustrating why the average treatment effect of access does not identify the optimal trust policy for every worker. Structured training should pair successful and failed examples and teach users to distinguish generation, retrieval and verified evidence. Trust can also be institutional rather than purely technological. Staff may trust a model that they

would otherwise doubt because it was approved by a prestigious vendor, endorsed by leadership or embedded inside familiar software. Conversely, strong technical performance may not overcome fears about surveillance, deskilling, discriminatory outcomes or job displacement. For organizational adoption to remain legitimate, governance must be intelligible and subject to meaningful employee voice. The relationship between trust and effective performance is likely to be non-linear. At very low trust, useful systems are ignored and complementary capability investments yield weak returns. At moderate, evidence-grounded trust, users employ models selectively, check claims and escalate high-risk outputs. At excessive trust, automation bias and verification fatigue allow failure rates to rise even when nominal adoption and satisfaction are high. A central research implication is that studies should measure reliance calibration against objective performance, not simply respondents’ favorable attitudes toward AI. Monitoring for calibrated trust requires behavioral indicators: appropriate acceptance rate, inappropriate acceptance rate, appropriate rejection rate, rate of escalation, error detection time and reliance behavior across task risk levels. Experimental test cases with known ground truth should complement attitudinal scales. Organizations should distinguish confidence in system competence from confidence in governance fairness and procedural legitimacy, since each may follow a different causal pathway.

Human-AI Collaboration and Responsible Work Redesign

GenAI contributes through complementary labor rather than automatic substitution. Tasks differ in codifiability, verifiability, creativity, accountability and access to trusted data. Routine drafting may admit relatively autonomous generation with sampling-based review; legal, medical, financial and employment decisions often require detailed human control. Vaccaro et al. (2024), in a large systematic review and meta-analysis of human–AI

collaboration experiments, show why combining humans and AI does not consistently outperform the better standalone actor. Complementarity is a design problem rather than a natural consequence of co-use. Responsible workflow redesign specifies the role of the machine at each decision stage: idea generator, evidence organizer, simulator, critic, preliminary triage system or constrained action executor. The human role is similarly specified: framing objectives, adjudicating trade-offs, validating evidence, authorizing consequential action and learning from outcomes. Unclear role boundaries create a responsibility gap when a fluent machine output is treated as a recommendation yet no person owns the final decision.

Knowledge-work processes demand special safeguards against epistemic displacement. Generative answers can obscure the provenance of claims or create artificial consensus. Alshammari (2026) analyzes tensions involving epistemic authority and knowledge governance in relation to innovation novelty and reliability. A robust organization preserves access to underlying source material, a channel for expert dissent and procedures to document departures from model advice. This is especially important where innovation prizes novelty and where rare failures carry disproportionate costs.

Skills policy is an integral part of governance. Training solely in prompting is inadequate without statistical literacy, privacy awareness, domain verification and the practical ability to override outputs. Managers need capabilities in portfolio selection, risk ownership, vendor negotiation and benefit realization. The World Economic Forum (2025) emphasizes continuing changes in workforce skill needs; organizations should therefore plan skill redeployment and opportunities for progression rather than treating productivity increases as an exclusive financial gain.

Employee participation influences both technical robustness and distributive outcomes. Frontline users observe failure modes invisible to model benchmarks and frequently create informal workarounds. Structured worker voice can expose inequities between experienced and inexperienced employees or between well-resourced and marginalized teams. Conversely, punitive monitoring may suppress incident reporting and encourage shadow AI. Balanced governance evaluates not just compliance percentages but whether employees report risks early and whether corrective action is credible.

Sustainable Innovation: Pathways and Trade-offs

GenAI may accelerate sustainable innovation through search, recombination and experimentation. Teams can synthesize environmental regulations, propose alternative product architectures, draft

supplier engagement materials, explore circular business models and support analysis of complex sustainability disclosures. These are enabling capabilities, not evidence that every AI-generated recommendation is environmentally beneficial. Innovation outcomes require independent measures of novelty, usefulness, implementation success and net social or ecological consequences.

Organization information processing theory provides a mechanism: environmental and regulatory challenges create uncertainty, equivocality and information loads, while GenAI can enhance interpretive capacity. Bag et al. (2026) report evidence linking GenAI integration, ESG sensemaking and environmental performance, while emphasizing that information overload may weaken value creation. The present framework extends this account by suggesting that trust calibration shapes whether interpretations are verified and translated into decisions rather than merely circulated. The innovation process can be parsed into opportunity identification, idea diversity, selection, development, implementation and diffusion. GenAI may add ideas at the first two stages, yet its statistical tendency to reproduce prevalent patterns may reduce the diversity of proposals. Doshi and Hauser (2024) show why individual creative benefits may coexist with collective-level trade-offs, while Meincke et al. (2025) highlight concerns around brainstorming diversity. Organizations should use explicit diversity prompts, independent human ideation, competing model outputs and blind review of options to prevent premature convergence.

Sustainable innovation should be defined through net effects. A generative design workflow that cuts engineering time but encourages unnecessarily frequent product turnover may have a negative ecological balance. Likewise, lower administrative labor costs may coincide with worker precarity if productivity gains are not accompanied by capability investment and fair transitions. Triple-bottom-line evaluation requires baseline comparisons, time horizons and distributional analysis rather than an assumed identity between digitalization and sustainability.

Empirical evidence is emerging but context dependent. Chen et al. (2026) examine stages of GenAI assimilation and sustainability performance; Feng et al. (2026) analyze AI-driven green governance using firm data, but their broad AI variable should not be conflated with GenAI alone. Bag et al. (2026) focus on ESG sensemaking, and studies of GenAI-enabled digital servitization link organizational mechanisms to sustainability outcomes. Collectively, they support further hypothesis testing, not a universal claim that AI adoption causes positive environmental change.

A dynamic capability perspective suggests a sequence: governance makes experimentation safe enough to learn; calibrated trust enables selective

reliance; organizational learning converts trials into validated routines; and environmental and social accounting supports responsible scaling. Managers should resist an innovation portfolio dominated by easily counted outputs prompts, drafts, generated

ideas and instead examine survival of innovations, reduction in material usage, quality of stakeholder outcomes and enduring economic value.

Table 4. Sustainable innovation mechanisms and countervailing risks

Innovation phase	GenAI opportunity	Sustainability or trust risk	Assurance response
Discovery	Faster synthesis of options	Popular-pattern convergence	Independent ideation and diversity testing
Design	Simulation and alternative proposals	False engineering claims	Domain review and source traceability
Pilot	Rapid testing and documentation	Unobserved social effects	Stakeholder review and pilot stop rules
Scale	Knowledge diffusion	Rebound demand and energy burden	Lifecycle and task-normalized accounting

Organizational Performance and Multidimensional Value

Organizational performance must be disentangled from isolated task productivity. A shorter time to complete a report may not improve organizational results if review burden increases, errors trigger rework or customers lose trust. Net performance therefore combines throughput, accuracy, rework, service outcomes, learning, resilience, compliance and resource costs. Field research indicating a productivity gain in one occupation is informative, but extrapolation across industries must account for task mix, worker expertise, integration costs and governance overhead (Brynjolfsson et al.,2025). Economic performance encompasses operating margins, revenue from new offerings, cycle times, process variability, quality-adjusted productivity and risk-adjusted returns. The costs to include are licensing, inference usage, specialist staff, data engineering, assurance, cybersecurity, vendor lock-in, training and failure remediation. An organization that reports gross time savings without deducting additional review time will systematically overstate value. Robust evaluation uses comparison groups and pre-specified metrics rather than success stories alone. Social performance includes employee well-being, fairness in work allocation, equal access to augmentation, customer inclusion, privacy, occupational learning and the transparency of consequential decisions. Generative tools can democratize expertise while simultaneously magnifying unequal skill and data access. AI-enabled HRM research indicates the relevance of trust to engagement and employee performance (Prasad & De,2024), but firm-level causal

attribution requires longitudinal controls and objective outcomes. Environmental performance involves electricity use, carbon intensity, water demand where relevant, hardware and infrastructure lifecycles, avoided material waste and the consequences of induced demand. The IEA (2025) identifies substantial interactions between AI demand and the energy system. A defensible environmental metric estimates net avoided emissions and resource use relative to a realistic baseline, with uncertainty intervals and clear boundaries. Claims about “green AI” should specify whether they concern the footprint of AI or AI-enabled reductions elsewhere. Performance also depends on resilience. Governance can initially increase task time by requiring verification, but this cost may be rational if it reduces severe incidents. Leon (2026) provides a salient illustration of this assurance–efficiency trade-off. Accordingly, an organizational dashboard should combine leading indicators of governance and trust with lagging performance outcomes rather than reward short-term usage growth as a success proxy. The performance function is therefore contingent and dynamic. Governance can be negatively associated with speed in the short run while positively associated with quality and risk-adjusted value over longer periods. A credible research design must observe this timing: immediate latency and compliance costs, medium-term process adaptation and long-run innovation or sustainability outcomes. The conceptual model predicts both direct and mediated relationships, not a uniformly positive linear effect in every setting.

Table 5. Balanced performance measurement framework

Dimension	Leading indicators	Lagging indicators
Economic	Validated use cases; unit task cost	Quality-adjusted productivity; net margin
Social	AI literacy; appeals access; inclusion	Employee outcomes; fairness disparities
Environmental	Energy per accepted outcome; model efficiency	Net lifecycle emissions; avoided material waste
Risk / resilience	Escalation success; tested controls	Incident loss, service reliability, audit findings
Innovation	Pilot quality; idea diversity	Adopted sustainable offerings; survival rate

Integrated Model and Testable Propositions

The model depicted in Figure 1 combines a governance antecedent, behavioral mediation, enacted human-AI collaboration and two downstream outcome domains. Governance maturity (GM) is multidimensional: structural assignment of accountability, procedural risk management, technical evaluation and relational legitimacy. Algorithmic trust calibration (ATC) measures how well reliance behavior matches task-specific system capability, not whether employees strongly agree that AI is trustworthy. Responsible AI use (RAU) captures verification, documented escalation, human authorization and learning routines.

Sustainable innovation capability (SIC) captures an organization's ability to turn validated AI-supported exploration into economically viable and environmentally or socially beneficial solutions. Organizational performance (OP) aggregates a transparent set of economic, social and environmental indicators while preserving separate domain analyses. The model includes direct GM-to-RAU and GM-to-OP pathways because governance may prevent losses independently of trust. It also treats task criticality, literacy, culture and regulatory intensity as moderators, not as background descriptors.

Proposition P1: Higher GenAI governance maturity is positively associated with employees' objectively measured trust calibration, conditional on perceived legitimacy of governance controls. Mechanism: clear scope of use, reliable performance evidence and understandable escalation procedures help employees identify when reliance is warranted. A high level of formal control without legitimacy can weaken the mechanism through circumvention.

Proposition P2: Trust calibration positively predicts responsible GenAI use and decision quality, whereas raw trust intensity exhibits an inverted-U relationship with decision quality in settings of uncertain model reliability. The latter proposition separates the motivational benefits of willingness to use a system from the dangers of over-reliance. It can be tested with seeded-error tasks and observed acceptance behavior.

Proposition P3: Responsible AI use mediates the association between governance maturity and sustainable innovation capability. Governance supports experimentation by limiting unacceptable risks; verified collaboration converts exploration into implementable products, processes or business models. Responsible use is expected to improve the quality and survival of innovations rather than necessarily maximize the number of generated ideas.

Proposition P4: Sustainable innovation capability mediates the long-term association between responsible AI use and environmental and economic performance. Short-term efficiency gains may precede this effect, but the expected innovation-mediated association should be stronger over longer observation periods and in firms with complementary data and change-management capabilities.

Proposition P5: Decision criticality strengthens the positive effect of risk-proportionate governance on error prevention while increasing verification costs. The resulting net economic relationship can be non-monotonic. Excessively heavy controls for trivial use cases discourage adoption, while insufficient controls for consequential tasks increase the probability of severe losses.

Proposition P6: Employee AI literacy positively moderates the effect of trust calibration on responsible use; participatory governance strengthens the association between formal controls and perceived legitimacy. These propositions recognize that technology adoption unfolds in social systems where employees interpret rules and negotiate the acceptable division of labor.

Proposition P7: Sustainability-oriented governance strengthens the association between innovation capability and net environmental performance by requiring lifecycle accounting, materiality thresholds and monitoring of rebound effects. When resource costs of model inference and induced demand exceed benefits, innovation output may increase while environmental performance declines.

Proposition P8: The interaction between governance maturity and trust calibration predicts stronger risk-adjusted organizational performance than either

construct alone. This complementarity hypothesis requires testing against alternative explanations: effective management may simultaneously produce good governance, trust and high performance

without governance itself causing the observed outcomes.

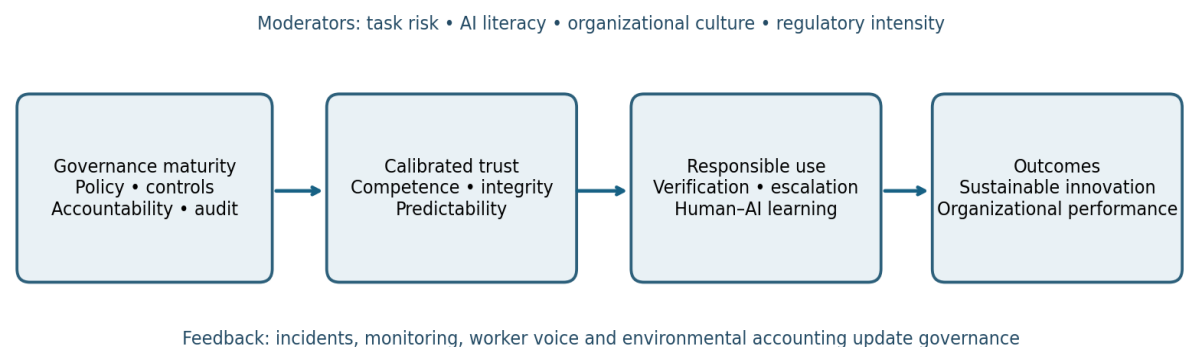


Figure 1. Proposed governance-trust-responsible use-performance framework (author-developed; hypothetical relationships).

Table 6. Propositions and suggested designs

Proposition	Expected relationship	Preferred validation
P1	Governance maturity → trust calibration (+), legitimacy contingent	Audits plus experimental reliance tasks
P2	Trust calibration → responsible use (+); raw trust → quality (inverted U)	Seeded-error behavior and longitudinal logs
P3	Responsible use mediates governance → innovation	Three-wave organizational panel
P4	Innovation mediates responsible use → long-run performance	Firm panel with sustainability outcomes
P5	Task risk moderates' governance-error and cost effects	Cross-risk randomized field design
P6	Literacy and participation amplify responsible use	Multilevel intervention
P7	Sustainability governance improves net environmental impacts	Lifecycle and counterfactual study
P8	Governance × calibration predicts risk-adjusted results	Firm-level longitudinal interaction analysis

Measurement and Empirical Research Agenda

An empirical extension should sample organizations at multiple levels of AI maturity across sectors with materially different risk profiles. A feasible design combines a baseline employee survey, independent governance audits, logged human-AI task evaluations and follow-up organizational performance measures at six and twelve months. Stratified sampling by industry, organization size and use-case criticality would reduce the risk of extrapolating from technology-forward firms alone. Governance maturity can be assessed by trained reviewers using documentary evidence: approved use-case register completeness, assigned owners, frequency of validation, effectiveness of escalation, third-party assessments, change control and incident closure. A 0-4 anchored rubric is preferable to management self-reports alone. The reviewer should

examine both whether a control exists and whether sampled cases demonstrate operational use. Inter-rater agreement and external audit validation are important. Trust calibration should be measured with paired self-report and behavioral tests. Participants complete tasks with known answer quality and variable error probability. For each trial researchers record confidence, acceptance, verification steps, final answer correctness, override decisions and time. Calibration can be computed as the discrepancy between reliance probability and model correctness across bins or as a decision-theoretic loss weighted by task severity. A single favorable-attitude scale is insufficient. Responsible use can be operationalized through the proportion of consequential decisions with recorded human approval, citation verification, correct

escalation, privacy-safe prompting and demonstrated response to model warnings. These indicators should be adjusted for use-case risk. Innovation capability can be measured by validated idea-to-pilot conversion, novel sustainability-oriented products, diffusion and independent assessment of beneficial novelty. Performance should be observed from organizational records rather than entirely inferred from respondent perceptions. Multilevel modeling is appropriate when employees are nested in teams and organizations; structural equation modeling may test latent constructs but does not establish causal effects from cross-sectional data. Longitudinal mediation requires temporal separation of antecedents, mediators and outcomes.

Causal identification strengthened by staggered governance rollouts, random assignment of explanatory controls or natural changes in vendor capabilities, provided ethical and legal constraints are satisfied. Robustness checks should include alternative trust metrics, pre-registration, objective benchmark tasks, organization fixed effects, source-separated measurement, sensitivity to missing data and checks for attrition. Environmental analyses should adopt explicit functional units and uncertainty ranges. Comparisons across firms require measurement invariance; language, industry regulation and cultural norms may affect what employees mean by “trust” or “accountability.” The literature warrants more design-based field research rather than further invalidated correlations alone.

Table 7. Draft survey and audit measurement items (illustrative, not validated)

Construct	Example measurement	Data source
Governance	“Every consequential GenAI use case has a documented owner.”	Independent document audit
Legitimacy	“I can challenge an AI-supported recommendation without penalty.”	Anonymous employee survey
Trust calibration	Difference between reliance frequency and test-case correctness	Behavioral experiments
Responsible use	Share of high-risk outputs receiving required independent review	Workflow events
Innovation	Share of pilots meeting novelty and verified impact criteria	Innovation portfolio review
Performance	Net value compared with preregistered non-AI baseline	Finance, HR and sustainability records

Illustrative specification (not estimated): $RAU_{ij,t+1} = \beta_0 + \beta_1 GM_{j,t} + \beta_2 ATC_{ij,t} + \beta_3 (GM_{j,t} \times ATC_{ij,t}) + \beta_4 X_{ij,t} + u_j + \varepsilon_{ij,t+1}$, where i indices’ employees, j organizations, t observation waves, X covariates, u organization-level heterogeneity and ε unexplained variation. A sequential model would then relate $SIC_{j,t+2}$ and $OP_{j,t+3}$ to earlier governance and responsible use. Coefficients are placeholders for future estimation; none are empirical results.

Managerial Implications and Implementation Roadmap

The practical recommendation is to begin with the work rather than with the model. Leaders should inventory decision processes, identify where GenAI can relieve information bottlenecks and classify the consequences of incorrect outputs. A portfolio approach helps distinguish approved productivity augmentation from novel applications requiring deeper assurance. Each use case should have a business sponsor, a risk owner, documented ground truth for evaluation and a defined stopping rule. Governance committees need cross-functional representation spanning operations, IT, data

protection, security, legal compliance, human resources, sustainability and employee stakeholders. Their remit includes prioritizing deployment, setting assurance depth, reviewing severe incidents and challenging assumptions about ROI. Incentives must not reward only the scale of usage; they should also value incident reporting, evidence quality and responsible cancellation of underperforming applications. Organizations should pilot with explicit baselines and predefined decision thresholds. For low-stakes tasks, automated format checks and sampled reviews may suffice. For medium-risk activities, source verification and second-person approval may be warranted. For high-impact decisions, domain-specific tests, accessible contestation, human decision authority and auditable logs become critical. The assignment of controls should follow context-specific risk rather than a universal one-size-fits-all checklist. Implementation requires operational feedback. An evaluation cadence should be triggered by changes in the model, prompt, workflow, retrieval corpus or affected population; by material incidents; and by scheduled risk review. Teams should maintain

“failure libraries” that preserve anonymized problematic outputs and train reviewers to identify error classes. Feedback must inform both technical changes and job design: some recurring errors are failures of organizational incentives, not only model capability. Finally, sustainability criteria must be included at the project gate. A new GenAI service should

document energy or cost proxies, expected avoided material or labor waste, social risks and rebound scenarios. The approval case should compare the project against a feasible non-GenAI alternative. The objective is sustainable net value, not merely deployment of an apparently advanced technology.



Figure 2. Iterative implementation roadmap for GenAI governance (author-developed).

Table 8. Stage-gate management actions and documentation

Gate	Required artifact	Example go/no-go criterion
G1: mandate	Use-case statement; stakeholders; baseline	Specific measurable problem, accountable sponsor
G2: risk	Privacy and impact assessment	Risk appetite and controls matched to task
G3: evaluation	Test set, errors, accessibility assessment	Prespecified error bounds satisfied
G4: deployment	Training, override rules, incident pathway	Users can verify and contest outputs
G5: renewal	Value and sustainability scorecard	Net benefits, acceptable residual risk

Discussion and Theoretical Contributions

The framework contributes to governance theory by recasting AI assurance as an organizational complement to innovation rather than a binary restriction on it. Safeguards can expand the space of permissible experimentation by creating confidence that failures are detectable and bounded. This proposition is explicitly contingent: if safeguards are poorly designed, their burdens can suppress useful exploration and induce covert use. The theoretical object is the quality and legitimacy of governance, not simply the number of policies adopted. The article contributes to algorithmic trust research by moving from trust level to calibration. Trust should be evaluated in relation to objective system reliability, task consequences and observable verification behavior. This shift addresses the possibility that higher trust raises adoption but lowers decision quality. It also separates trust in the model from confidence in organizational accountability and procedural fairness, which can differ sharply. The socio-technical contribution is a process account linking policy to practice. Formal frameworks do not directly produce environmental outcomes; they shape the conditions under which workers experiment, rely, verify, contest and learn. Responsible use and innovation capability provide intervening mechanisms. This process view is consistent with empirical evidence showing heterogeneity in human-AI performance and with governance studies emphasizing implementation rather than principles alone.

The sustainability contribution is a strict distinction between AI-enabled sustainability opportunities and the full lifecycle footprint of AI. A reduction in drafting time or improved ESG narratives does not necessarily reflect lower emissions, better labor conditions or genuinely innovative products. Triple-bottom-line measurement and counterfactual reasoning are therefore necessary to prevent symbolic sustainability claims from substituting for real outcomes. The model is also compatible with negative cases. In a heavily regulated or safety-critical setting, governance may justifiably limit system autonomy and depress immediate output counts. In a low-risk creative setting, too much procedural oversight may be counterproductive. In weak data environments, fluent answers may increase false confidence without improving decisions. These exceptions are crucial to a theory intended to guide responsible practice across heterogeneous smart organizations.

Limitations and Directions for Future Research

The article is an integrative conceptual synthesis, not an original empirical test. Its proposed pathways should not be read as estimated causal effects. Search coverage is purposeful and does not meet the standards of a preregistered systematic literature review. The rapidly evolving field also means that the applicability of particular systems, regulations, prices or governance mechanisms may change. The evidence base overrepresents digitally intensive organizations and better-resourced settings.

Construct measurement remains challenging. Algorithmic trust and legitimacy are culturally and task dependent; sustainability outcomes unfold over long periods; and reporting practices can obscure resource consumption or displaced harm. External generalizability may be limited when studies focus on specific countries, industries, occupations or versions of models. Future work should combine cross-national comparative studies, objective behavior traces, longitudinal designs and independent sustainability accounting.

Promising research questions include the optimal amount of explanation for different tasks, the causal effect of employee participation on circumvention, how governance influences diffusion of beneficial novelty, when small or domain-specific models outperform larger models on net value, and whether sustainable innovation mediates financial gains after full cost accounting. Research should explicitly report negative outcomes, decommissioned projects and cases where conventional technology proves superior.

Further work could test nonlinear governance effects, distinguish workplace trust from external stakeholder trust, investigate power asymmetries in AI-supported decision making and model changes in the meaning of human accountability as agentic capabilities develop. Such research will determine which components of the conceptual framework travel across institutional and technological settings.

Conclusion

Generative AI governance and algorithmic trust should be treated as mutually reinforcing organizational capabilities rather than separate matters of compliance and user sentiment. The integrative framework proposes that risk-proportionate, participatory and auditable governance supports calibrated reliance, which enables responsible human-AI work practices and potentially stronger sustainable innovation and organizational performance. However, positive outcomes depend on task design, employee literacy, managerial legitimacy, environmental externalities and the quality of implementation. The central managerial implication is to pursue verified net value, not maximum AI use or maximum trust.

Supplementary Analysis I. Competing Causal Explanations

An important rival to the governance-first account is resource-based selection. Better-managed firms typically have superior information systems, larger compliance teams, stronger innovation portfolios and more attractive employment conditions. These underlying advantages may simultaneously predict governance maturity, employee confidence and economic performance. Consequently, finding a positive correlation between governance scores and profitability would not isolate a causal governance

effect. Future research should collect pretreatment measures of digital maturity, management quality, organizational size, prior innovation, data accessibility and financial slack. Fixed-effects models can remove stable organizational heterogeneity, but time-varying confounding and anticipatory adoption remain important threats.

A second rival explanation is institutional conformity. Firms may document sophisticated AI policies because clients, investors or regulators demand visible safeguards, without substantially changing employee workflows. Under this decoupling hypothesis, governance predicts favorable audit scores but has little relationship with actual error detection, escalation or contestability. A discriminating test would randomly audit a sample of model-assisted decisions and match them to contemporaneous compliance documents. The gap between stated policy and verified practice should be a distinct measured construct. In principle, formal governance can even lower trust if users infer that elaborate safeguards signal an especially hazardous system. A third explanation concerns reverse causality: successful AI applications might produce resources and executive attention that fund stronger governance, rather than the other way around. This is particularly plausible when management sponsors first deploy a tool experimentally and add structures only after a visible success or incident. Researchers should record adoption sequences, governance investments and work redesign milestones prospectively. A difference-in-differences design around staggered policy introductions may be useful when the parallel-trends assumption can be defended and adoption timing is not itself driven by anticipated changes in performance. Fourth, apparent trust effects may conceal expertise effects. A capable employee may both detect inaccurate outputs and express appropriately conditional trust, while generating better work outcomes independently of GenAI. In this case, calibration is partly a behavioral marker of expertise, not necessarily the principal causal mediator. A factorial experiment crossing AI training, disclosure quality and independent verification access would help distinguish changes in attitudes from changes in skills. A particularly useful dependent variable is correct final task performance after accounting for the amount of human effort and consultation time.

Finally, the model should accommodate failure and substitution. In domains where externalities are extreme or evidence is poor, eliminating a GenAI application may be the most effective governance decision. Responsible decommissioning and nonadoption should count as governance successes when they improve counterfactual risk-adjusted value. An innovation system assessed solely by number of deployed assistants risks biasing conclusions toward expansion even when restraint would better advance sustainability. This possibility

distinguishes the proposed account from technologically deterministic models in which every increase in digital intensity is assumed to imply progress.

Supplementary Analysis II. Trust Calibration and Evaluation Protocol

A pragmatic trust-calibration experiment would recruit employees who perform comparable tasks and randomize them to several interface or governance conditions. Participants would first solve an independent baseline set without AI, establishing subject-matter competence. They would then complete balanced task sets with a model whose correctness has been independently adjudicated by domain experts. The experimental conditions might differ in access to source evidence, presentation of limitations, user training and escalation options. Tasks should span low and high consequences and include both straightforward and deceptive outputs to prevent participants from equating fluency with reliability. A complete trial record would contain the original task, model version, prompt, context sources, generated suggestion, objective correctness label, employee confidence rating, decision to adopt or reject, verification actions, decision latency and final outcome. Privacy controls must prohibit collecting unnecessary personal or confidential data, and organizational participants should not be penalized for experimentally induced errors. Independent experts should adjudicate disputed answers without knowing the experimental condition. Balanced difficulty across arms and preregistered exclusions would protect the validity of group comparisons. Several complementary calibration statistics should be considered. The simplest compares acceptance rates on correct and incorrect outputs; appropriate reliance requires accepting more reliable advice while rejecting misleading advice. A signal-detection framing can separate discrimination from the general tendency to accept automation. Probability calibration can compare numerical confidence with realized correctness when respondents provide calibrated probabilities, whereas expected decision loss can attach higher penalties to errors in high-impact tasks. Researchers should report these quantities by occupational group, experience level and decision risk rather than rely on a single pooled trust score.

Evaluation should examine the behavioral cost of controls. For instance, mandatory human review may reduce incorrect acceptances but increase average resolution time and reviewer fatigue. The appropriate choice depends on the severity and frequency of failures and the availability of more efficient safeguards, including selective review, source grounding, constrained outputs, specialist models or abstention rules. Comparative effectiveness must be measured against a

meaningful no-AI or alternative-technology baseline. A system that looks safe in isolation may still produce worse decisions than a well-designed traditional workflow. The ethical dimension of experimentation matters. Employees need to know the general purpose of assessment, the scope of logging, data retention periods and whether managers can inspect individual behavior. Where deception is unavoidable for valid error detection, ethical approval and appropriate debriefing are necessary. A separate anonymous channel should capture participants' perceptions of fairness, psychological safety and pressure to rely on the system. The research question is not simply whether a safety control works technically, but whether it remains effective when introduced into an actual work environment with competing incentives.

Supplementary Analysis III. Sustainability Accounting and Reporting

Attribution is particularly difficult for environmental claims. A GenAI system may support a product redesign that uses less material, but the associated benefit also depends on manufacturing changes, supplier behavior, customer demand and product lifetimes. Environmental claims should therefore specify a counterfactual baseline, functional unit, temporal boundary and scope of included processes. The analysis should include additional computing and storage, data transfer where material, validation and human review. Estimated impacts should be expressed as ranges when model-provider energy information or supplier emissions data are incomplete.

Three complementary assessment levels are useful. At the task level, organizations can estimate energy, cost and accepted-output quality for competing model configurations. At the process level, they can compare full human-AI workflow resource consumption with the existing workflow, including rework and the possibility of additional output demand. At the organizational level, they can examine whether productivity gains shift rather than reduce total resource use. A narrow per-query efficiency metric is insufficient to establish sustainability when lower unit cost greatly expands the volume of queries.

Social sustainability requires comparable care. Workforce savings should not automatically be reported as welfare gains; outcomes depend on whether employees experience greater autonomy, skill development, safe workloads, equitable access and meaningful participation. A model that benefits average productivity while concentrating errors or exclusion in a minority group may perform poorly on distributional criteria. Accordingly, decision dashboards should report both averages and disaggregated outcomes. Stakeholder appeals, worker perceptions of legitimacy and customer

complaint patterns are useful complements to quantitative measures of technical accuracy. For senior executives, the reporting template should link each use case to its owner, risk tier, baseline, latest quality-adjusted benefit, direct operating cost, documented incidents, review coverage, environmental indicators and date of reassessment. Portfolio summaries should distinguish exploratory projects, approved production tools, restricted applications and retired systems. This creates an audit trail connecting strategy to implementation. It also permits organizations to report honestly when sustainability impacts are uncertain or when technological alternatives outperform a generative solution. The long-run research opportunity is to compare alternative governance portfolios rather than isolated controls. Different bundles of assurance, worker training, model selection, data architecture and sustainability review may yield comparable error prevention at very different economic and environmental costs. Comparative studies can use matched organizations, intervention sequences or configurational methods to identify contextually effective combinations. Such analysis would advance a contingencies-based theory in which governance is valued for enabling responsible and durable organizational outcomes, not for satisfying uniform procedural checklists.

Supplementary Analysis IV. Governance in Organizational Scenarios

Consider an internal knowledge assistant in a medium-sized manufacturing firm. It retrieves approved engineering manuals and helps maintenance staff locate troubleshooting procedures. The system appears low risk when it summarizes general training content, but the risk rises sharply if staff interpret answers as authorization to override equipment safety controls. Governance should therefore specify permitted informational uses, block operational commands outside approved processes and require explicit source citations. Trust calibration tests can seed incorrect maintenance statements and evaluate whether employees consult the underlying documentation. Performance evaluation should count accurate resolution, time spent verifying instructions and incidents avoided, rather than measuring response speed alone.

A second scenario concerns an organization using GenAI to draft sustainability disclosures. The assistant might reduce transcription costs and assemble evidence from disparate departments, but it can also invent environmental achievements or omit unfavorable material facts. Controls should preserve traceability from each assertion to validated source records, require sign-off by accountable sustainability and financial specialists and keep a revision history. A successful pilot would demonstrate fewer reconciliation errors, improved

evidence coverage and reduced administrative work without increasing misleading statements. Sustainability benefits should be evaluated against the additional computational work and any resulting improvements in actual management decisions.

A third scenario involves customer service in a multilingual service provider. GenAI may extend coverage, help less experienced employees and improve service consistency. However, differential answer quality across languages can create inequities even as average satisfaction rises. The governance design should therefore use representative test sets, regionally appropriate escalation channels, clear advice boundaries and monitoring of complaints by user group. Trust should be examined separately for employees who use suggestions and customers who experience the resulting service. A high user satisfaction score cannot substitute for verified procedural fairness, privacy protection or contractual accuracy.

These scenarios suggest a general principle: the same foundation model can generate different governance obligations because consequences, actors, evidence and downstream authority differ. Comparative empirical work should categorize applications by decision effect rather than by vendor or model name. A useful research unit is the governed use-case workflow, nested within organizational capabilities. This enables researchers to observe policy alignment, actual human behavior and measurable outcomes while avoiding the mistaken assumption that company-wide AI adoption is a homogeneous treatment.

Declarations

Funding: No external funding is claimed for this manuscript. Conflicts of interest: No conflicts of interest are declared. Data availability: No primary dataset was created or analyzed. Ethics approval: Not applicable to this non-empirical conceptual review. Generative AI disclosure: This manuscript was drafted with AI assistance; all bibliographic records and article-specific claims should be checked again against publisher versions before journal submission. Authorship: No person or institutional affiliation is attributed without authorization.

Acknowledgments

The authors would like to express their sincere gratitude to Ravan Gostaran Tavan Gharb Company for its invaluable support, effective cooperation, and generous assistance throughout the preparation and development of this manuscript. The authors gratefully acknowledge the company's contribution to the successful completion of this work.

Disclosure Statement

No potential conflict of interest reported by the authors.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Authors' Contributions

All authors contributed to data analysis, drafting, and revising of the paper and agreed to be responsible for all the aspects of this work.

References

- [1] Alshammari, K. H. (2026). [Generative AI, epistemic authority, and knowledge governance tension: Implications for innovation novelty and reliability](#). *Future Business Journal*, 12, Article 222.
- [2] Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). [Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile \(NIST AI 600-1\)](#). *National Institute of Standards and Technology*.
- [3] Bag, S., Srivastava, G., Routray, S., & Chiarini, A. (2026). [Generative AI, ESG sensemaking, and environmental performance: An OIPT perspective](#). *Business Strategy and the Environment*, 35(5), 7196-7217.
- [4] Batool, A., Zowghi, D., & Bano, M. (2025). [AI governance: A systematic literature review](#). *AI and Ethics*, 5, 3265-3279.
- [5] Brynjolfsson, E., Li, D., & Raymond, L. (2025). [Generative AI at work](#). *The Quarterly Journal of Economics*, 140(2), 889-942.
- [6] Burton, J. W., et al. (2024). [How large language models can reshape collective intelligence](#). *Nature Human Behaviour*.
- [7] Chen, J., Tian, T., Wong, C. W. Y., & Lai, K.-H. (2026). [How stage-based generative artificial intelligence \(GAI\) assimilation impacts sustainability performance: The role of the technology-organization-environment \(TOE\) framework](#). *Sustainable Futures*, 11, 101918.
- [8] Doshi, A. R., & Hauser, O. P. (2024). [Generative AI enhances individual creativity but reduces the collective diversity of novel content](#). *Science Advances*, 10(28), eadn5290.
- [9] Feng, B., Chen, X., & Tang, H. (2026). [AI-driven green governance: Assessing the impact of artificial intelligence on corporate sustainability performance](#). *Journal of Innovation & Knowledge*, 11, 100869.
- [10] Janssen, M. (2025). [Responsible governance of generative AI: Conceptualizing GenAI as complex adaptive systems](#). *Policy and Society*, 44(1), 38-51.
- [11] Lee, B. C., & Chung, J. (2024). [An empirical investigation of the impact of ChatGPT on creativity](#). *Nature Human Behaviour*, 8, 1906-1914.
- [12] Leon, M. (2026). [Governing generative AI in organizations: A design theory and quasi-experimental field study of sociotechnical guardrails](#). *The Journal of Supercomputing*, 82, Article 641.
- [13] Luengo Vera, C., Bhimani, A., Gómez Gandía, J. A., & de Lucas, A. (2026). [Generative AI and the algorithmic workplace: A bibliometric and conceptual analysis of its impact on organizational decision-making and work design](#). *Journal of Open Innovation: Technology, Market, and Complexity*, 12(1), 100710.
- [14] Meincke, L., Nave, G., & Triesch, C. (2025). [ChatGPT decreases idea diversity in brainstorming](#). *Nature Human Behaviour*, 9, 1107-1109.
- [15] Organization for Economic Co-operation and Development. (2024). [Explanatory memorandum on the updated OECD definition of an AI system \(OECD AI Papers No. 8\)](#).
- [16] Prasad, K. D. V., & De, T. (2024). [Generative AI as a catalyst for HRM practices: Mediating effects of trust](#). *Humanities and Social Sciences Communications*, 11, Article 1362.
- [17] Schilke, O., & Reimann, M. (2025). [The transparency dilemma: How AI disclosure erodes trust](#). *Organizational Behavior and Human Decision Processes*, 188, 104405.
- [18] Stanford Institute for Human-Centered Artificial Intelligence. (2026). [The AI Index Report 2026: Economy](#). *Stanford University*.
- [19] Sullivan, V., & Weger, K. (2025). [Transparency and explainability in AI-assisted decision making: Effects on trust, perceived reliability, confidence, and ease of understanding](#). *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 69(1).
- [20] Vaccaro, M., Almaatouq, A., & Malone, T. (2024). [When combinations of humans and AI are useful: A systematic review and meta-analysis](#). *Nature Human Behaviour*, 8, 2293-2303.
- [21] World Economic Forum. (2025). [The Future of Jobs Report 2025](#).
- [22] Zaheer, A., Subramaniam, Y., Zayed, N. M., Bipasha, M. S., & Alrawi, W. K. (2026). [Generative AI utilization and employee productivity through digital skills with the moderating role of AI trust](#). *Discover Sustainability*.