

Review

From Copilots to Coworkers: How Autonomous AI Agents Reshape Organizational Roles and Decision Authority: A Structured Integrative Review and Research Agenda

Miguel Rozalez^{1*}

¹Hillsborough Community College, United States.

Keywords:

Autonomous AI Agents;
Agentic AI;
Human–AI Collaboration;
Organizational Design;
Decision Authority;
Algorithmic Management;
Accountability;
Future of Work

ARTICLE INFO

ABSTRACT

Autonomous artificial intelligence agents extend generative AI from producing content on demand to pursuing goals through planning, memory, tool use, environmental feedback, and iterative action. This shift matters organizationally because agency is no longer confined to recommendation: systems can increasingly initiate and coordinate consequential work. This article integrates research on AI-enabled work, algorithmic management, organizational design, human – AI collaboration, automation, and responsible AI to explain the transition from copilots to digital coworkers. The review distinguishes five configurations—tool, copilot, teammate, delegate, and coworker—and

argues that the central organizational variable is not technical capability alone but the allocation of decision authority. We synthesize evidence showing that AI assistance can improve speed and quality in bounded tasks, often benefiting less experienced workers, while also producing uneven performance at the frontier of model capability. Extending this evidence to agentic systems, we develop an authority allocation matrix based on task risk, reversibility, verifiability, and auditability. We identify three mechanisms through which agents reshape organizations: role reconfiguration, authority redistribution, and coordination redesign. Benefits arise when agents compress search and execution costs, diffuse expertise, and enable new forms of modular organizing. Risks arise when action outpaces verification, responsibility becomes diffused, tacit knowledge erodes, or agents acquire excessive access to organizational resources. We propose a layered governance architecture combining explicit mandates, least-privilege access, execution limits, provenance, continuous monitoring, and human escalation rights. The article concludes with propositions and a research agenda for studying agentic organizing through field experiments, longitudinal process research, digital trace analysis, and comparative institutional designs.

1 Introduction

Generative AI entered organizations primarily as a conversational interface: a user asked for a draft, summary, forecast, or recommendation, and the model returned an output. The organizational metaphor was therefore the copilot—a capable assistant that remained subordinate to a human operator. Autonomous AI agents alter that

*Corresponding author
Email: mrozalez@hawkmill.hccfl.edu

Online ISSN: 3136-4578

DOI: <https://doi.org/10.68471/fbai.2026.008>

Copyright: © 2026 The Author(s). Published by Global Frontiers Press Limited. This article is open access under the Creative Commons Attribution 4.0 International license (CC BY 4.0): <https://creativecommons.org/licenses/by/4.0/>.

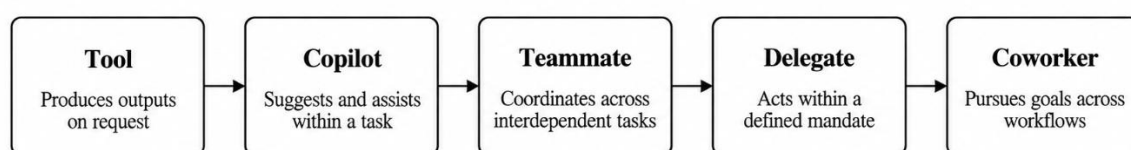
relationship. By combining foundation models with persistent memory, planning routines, external tools, application programming interfaces, and feedback loops, agents can decompose goals, select actions, inspect intermediate results, revise plans, and continue operating with reduced moment-to-moment supervision (Wang et al., 2024). The difference is not merely greater intelligence. It is the capacity to act over time and across organizational systems.

This transition moves AI from the informational layer of organizations toward the operational layer. A copilot may recommend which supplier to contact; an agent may identify suppliers, request quotations, compare terms, update a procurement system, and route an exception for approval. A copilot may draft a customer response; an agent may classify the case, retrieve account history, issue a permitted refund, document the action, and schedule follow-up. Each additional action transfers some discretion from a person to a technical system. Consequently, the defining management question is changing from “How should employees use AI?” to “Which decisions may an AI agent make, under whose authority, and with what means of contestation?”

Early productivity evidence provides a valuable but incomplete foundation. In a study of customer-support agents, access to a generative AI assistant increased issues resolved per hour and delivered the largest gains to less experienced workers (Brynjolfsson et al., 2025). An experiment with professional writing tasks found substantial reductions in completion time and improvements in output quality (Noy & Zhang, 2023). Field experiments with consultants revealed a jagged technological frontier: workers performed better on tasks within the model’s capabilities but could become less accurate when confident AI outputs crossed that frontier (Dell’Acqua, McFowland, et al., 2026). These studies concern assistance more than autonomous execution. Agentic systems intensify their implications because erroneous output can become erroneous action before a person inspects it.

The management literature has rich accounts of delegation, role design, routines, sociomateriality, and algorithmic control, but these streams have rarely been integrated around autonomous agents. Technical agent research, meanwhile, concentrates on architectures and benchmarks, often treating the organization as a deployment environment rather than a system of authority and accountability. This review bridges those conversations. It develops a role continuum, explains how decision rights migrate as autonomy grows, identifies organizational mechanisms and boundary conditions, and specifies a research agenda that can distinguish genuine transformation from vendor rhetoric.

The article makes four contributions. First, it offers a management-centered definition of an organizational AI agent. Second, it reframes agent adoption as a problem of decision-right allocation rather than a simple technology-acceptance problem. Third, it proposes a risk-sensitive authority matrix and an integrated framework linking antecedents, mechanisms, governance, and outcomes. Fourth, it sets out testable propositions and empirically tractable research questions for organization studies, information systems, strategy, human resource management, and business ethics.



The critical transition is from recommending actions to executing consequential actions.

Figure 1. From tool to coworker: a continuum of organizational AI agency.

2 Review Method and Conceptual Framework

2.1 Review design and scope

Because agentic AI is an emerging phenomenon whose vocabulary remains unstable, this study uses a structured integrative review rather than claiming the closure expected of a mature-domain meta-analysis. The review combines conceptually relevant work from management and information systems with recent technical and empirical research. The unit of analysis is a publication that contributes evidence or theory concerning at least one of four elements: autonomous or generative AI capability; human–AI task allocation; organizational roles and authority; or governance, accountability, and control.

Searches conducted through September 2026 used combinations of the terms “autonomous AI agent,” “agentic AI,” “large language model agent,” “human–AI collaboration,” “algorithmic management,” “decision authority,” “delegation,” “organizational design,” “AI productivity,” “human oversight,” and “AI accountability.” Priority was given to peer-reviewed journal articles, major conference proceedings, institutional working papers with transparent methods, and authoritative governance frameworks. Backward and forward conceptual chaining was used to connect contemporary agent research with established theories of automation, delegation, role systems, routines, and sociomateriality. Practitioner publications were used only when they documented an emerging organizational configuration unavailable in peer-reviewed work; they were not treated as causal evidence.

2.2 Analytical coding and synthesis approach

Each source was coded along six dimensions: system capability, autonomy level, task domain, human role, allocation of decision authority, and reported outcome or risk. A second-cycle synthesis grouped findings into three organizational mechanisms—role reconfiguration, authority redistribution, and coordination redesign—and four boundary conditions—task verifiability, reversibility, consequence severity, and auditability. This approach is theory-building rather than effect-size aggregation. The rapid pace of technical change also means that capability claims should be read as time-bound. The framework therefore emphasizes stable organizational design variables rather than predictions about a specific model.

Table 1. Evidence streams integrated in the review.

Evidence stream	Representative focus	Contribution to this review	Principal limitation
Workplace experiments	Productivity, quality, learning	Establishes heterogeneous effects of AI assistance	Mostly bounded tasks and copilot configurations
Agent architecture research	Planning, memory, tools, reflection	Defines how autonomous action becomes possible	Benchmarks may not represent organizational settings
Algorithmic management	Control, monitoring, worker autonomy	Explains power and role consequences	Often studies platforms rather than generative agents
Organization theory	Delegation, routines, decision rights	Provides mechanisms and boundary conditions	Predates adaptive foundation-model systems
Responsible AI and regulation	Oversight, transparency, risk controls	Informs governance architecture	Principles may remain abstract without process design

2.3 Evidence boundaries and source triangulation

The distinction between technical feasibility and organizational effectiveness is pivotal. Agent benchmarks assess whether a system can complete tasks under defined conditions, whereas organization-level claims additionally require evidence about authority, institutional context, implementation cost, and post-deployment outcomes. AgentBench, for example, provides useful evidence on the breadth and limitations of model-based agency, but benchmark success cannot establish that equivalent discretion should be granted in procurement, employment or regulated services (Liu et al., 2023). Accordingly, we triangulate technical studies with empirical workplace findings and organizational theory instead of assigning equal causal weight to every evidence stream.

This integrative synthesis has limits. Search coverage and source selection were guided by theoretical relevance rather than exhaustive systematic screening, and the review does not report a reproducible PRISMA search flow, formal risk-of-bias ratings or a pooled effect estimate. Propositions developed below should therefore be interpreted as analytical hypotheses and directions for validation, not as established causal findings. Future reviews could preregister inclusion criteria and code evidence quality separately for assistance and agentic execution.

3 From AI Assistance to Agentic Organizing

3.1 Defining organizational AI agents

An organizational AI agent is a software-based actor that receives an objective, maintains relevant state, selects and sequences actions, uses organizational tools or data, evaluates intermediate outcomes, and adapts its behavior within a delegated mandate. This definition excludes a conventional chatbot that merely responds to a prompt and a

deterministic workflow that follows fixed rules without adaptive planning. It also avoids attributing human consciousness or moral personhood to software. “Actor” refers to observable participation in an organizational process, not equivalence with a human employee.

Five capabilities jointly matter. Goal orientation allows the system to optimize toward an outcome rather than produce a single response. Planning decomposes the goal into steps. Memory carries information across interactions. Tool use connects reasoning to consequential systems such as email, customer relationship management, code repositories, financial software, or procurement platforms. Feedback and reflection allow the agent to compare results with objectives and revise its course. Autonomy is therefore multidimensional: an agent may be highly independent in planning but tightly constrained in execution, or may execute routine transactions freely while lacking permission to redefine the goal.

3.2 Why the copilot metaphor becomes insufficient

The copilot metaphor assumes a stable human principal who observes the work and retains immediate control. In practice, organizations seek value by reducing the attention required per task. When an employee must verify every intermediate action, automation may accelerate production but not coordination. Agentic systems promise to absorb the coordination burden by operating asynchronously, interacting with other agents and applications, and escalating only exceptions. The promise is economically important precisely because it relaxes the one-human-one-session constraint.

Yet reduced supervision changes the nature of risk. A flawed recommendation can be rejected; a flawed action can alter records, communicate externally, spend money, disclose data, or trigger other systems. Autonomy also creates temporal and relational distance between the person who configured a goal and the actions later taken in its name. The appropriate organizational metaphor is therefore neither employee nor tool alone. An agent is a delegated technical actor embedded in a chain of authority.

3.3 A continuum of organizational AI agency

Table 2. Organizational role configurations for AI systems.

Configuration	Primary function	Human role	Agent discretion	Illustrative use
Tool	Generate or transform an output	Direct operator	None beyond the prompt	Draft a market summary
Copilot	Recommend and assist within a task	Reviewer and decision maker	Local choice over content	Suggest responses to customers
Teammate	Coordinate an interdependent contribution	Collaborator and integrator	Select methods and intermediate steps	Analyze competitors while humans develop strategy
Delegate	Complete a bounded process	Mandate setter and exception handler	Execute within rules and thresholds	Resolve eligible service cases
Coworker	Pursue recurring goals across workflows	Supervisor, peer, and accountable owner	Replan, coordinate, and initiate permitted actions	Manage a campaign across analytics, content, and scheduling

The categories are analytical rather than chronological. The same technical system may act as a copilot in hiring, a delegate in inventory replenishment, and a tool in legal review. Classification should therefore follow the authority granted in a particular process, not the vendor’s product label. This process-level view also prevents organizations from treating autonomy as a binary attribute.

3.4 Evidence from current AI productivity research

The empirical literature offers three robust signals and one major gap. First, AI assistance can improve performance on bounded knowledge tasks. [Noy and Zhang \(2023\)](#) found faster completion and higher assessed quality in professional writing, while [Brynjolfsson et al. \(2025\)](#) documented productivity gains in operational customer service. Second, benefits are heterogeneous. Less experienced workers often gain more because AI makes codified patterns immediately available, but this does not imply that expertise becomes irrelevant. Experts remain better positioned to

recognize missing context, diagnose failures, and integrate competing considerations. Third, the performance frontier is uneven. Dell’Acqua, McFowland, et al. (2026) showed that generative AI can improve work inside its capability frontier while inducing confident errors outside it. The central gap is that most causal studies examine systems that advise or generate rather than systems authorized to act.

Team-level experiments also complicate simple substitution claims. The Cybernetic Teammate study suggests that individuals working with generative AI can approach or exceed the performance of conventional teams on some product-innovation tasks, while AI-assisted teams may achieve both stronger performance and more balanced functional contributions (Dell’Acqua, Ayoubi, et al., 2026). These results show that AI can change the value of team composition, but they do not demonstrate that an autonomous agent can replace the relational, political, and accountability functions of a team. Task performance is not organizational membership.

For autonomous agents, current evidence is dominated by technical benchmarks, demonstrations, and early deployments. Benchmarks establish that agents can browse, code, call tools, and complete multi-step tasks, but organizational validity is uncertain. Benchmark environments usually define goals and success criteria more clearly than workplaces do; they rarely capture disputed objectives, legacy-system constraints, tacit knowledge, reputational consequences, or the need to justify decisions to affected people. High benchmark scores should therefore be treated as evidence of capability under specified conditions, not a warrant for broad organizational authority.

The distinction between potential and realized autonomy is also important. A technically capable agent may operate under strict permissions and remain a copilot in practice. Conversely, a modest model connected to powerful tools may possess consequential execution authority. Empirical research should report both: what the system could do technically and what it was institutionally permitted to do. Without that distinction, studies risk conflating model capability, workflow design, and governance.

3.5 Agent architectures, simulated collaboration, and organizational transfer

Agent frameworks illustrate why a new work design vocabulary is needed. ReAct integrates reasoning traces with action selection, making sequential tool use technically tractable (Yao et al., 2023). Generative-agent simulations combine memory, reflection and planning to produce believable interactions in controlled environments (Park et al., 2023). AutoGen demonstrates how specialized language-model agents may exchange messages and share task responsibilities (Wu et al., 2024). These studies establish architectural possibilities, not independent evidence of labor substitution, managerial legitimacy or long-run productivity gains.

The organizational transfer problem concerns the gap between a successful demonstration and a dependable institutional role. In a laboratory scenario, objectives, permissions and termination criteria can be specified by researchers. In an enterprise, a task may be reinterpreted by customers, policies may conflict, and downstream systems may propagate errors. A deployment study should therefore distinguish capability validation, workflow validation, and governance validation. Only the third level can test whether human accountability and stakeholder recourse survive delegated execution.

4 Organizational Transformation through AI Agents

4.1 Task restructuring and role transformation

Delegation creates value when a principal transfers tasks to an actor with relevant capability, while retaining mechanisms to align behavior and observe outcomes. Traditional agency theory assumes human agents with interests that may diverge from those of principals. AI agents do not possess interests in the ordinary sense, but analogous alignment problems arise from objective misspecification, incomplete context, model behavior, vendor incentives, data bias, and interactions among tools. The key shift is from motivational opportunism to computational and institutional misalignment. Monitoring remains necessary, but monitoring every action defeats the economic purpose of delegation (Baird & Maruping, 2021).

This tension suggests that decision authority should be bundled with verifiability. Delegation is attractive where outputs can be checked cheaply, consequences are reversible, and exceptions can be detected. Where quality is ambiguous, effects are delayed, or actions affect fundamental rights, human authority must remain substantive rather

than ceremonial. The organizational problem is thus to design mandates that preserve the benefits of local autonomy without severing accountability.

This role-based interpretation is consistent with [Faraj et al. \(2018\)](#), who argue that learning algorithms can reshape expertise, occupational boundaries, and coordination. Agentic organizing extends these questions to systems that execute actions as well as generate information.

4.2 Expertise, learning and capability development

Organizations allocate information and decision rights to cope with uncertainty ([Simon, 1947](#); [Jensen & Meckling, 2001](#)). AI agents increase local information-processing capacity by searching, synthesizing, and acting at machine speed. This may decentralize decisions because more analysis can occur close to operational data. Paradoxically, it may also centralize control: a small group can encode objectives and policies into agents used across the enterprise. Agent adoption therefore need not flatten hierarchy. It can create a thin human hierarchy supported by centralized technical rules, or a more distributed structure in which teams configure specialized agents ([Shrestha et al., 2019](#)).

Whether agents substitute for middle management depends on what middle managers actually do. Reporting, scheduling, coordination, and policy enforcement are comparatively codifiable. Sensemaking, conflict resolution, political negotiation, identity work, and moral judgment are less so. Agents may eliminate some information-broker tasks while increasing demand for boundary setting, exception adjudication, model stewardship, and cross-functional governance.

[Jarrahi \(2018\)](#) locates human–AI complementarity in the combination of computational analysis with contextual and intuitive judgment. The implication for agents is that delegation should preserve access to human expertise when uncertainty cannot be reduced to a well-specified prediction task.

4.3 Identity, status and relational work

Roles emerge through repeated interaction between people, technologies, rules, and artifacts ([Orlikowski, 2007](#); [Zammuto et al., 2007](#)). An agent inserted into a workflow does not simply occupy a pre-existing job box; it changes what other participants attend to and how routines are performed. When an agent prepares a decision dossier, the manager’s work may shift from evidence gathering to validation. When it executes standard cases, frontline employees may receive a smaller but more difficult set of exceptions. Performance measures, training, and career paths must change accordingly ([Pachidi et al., 2021](#)).

Routines also supply organizational memory. Agents can externalize and distribute expert patterns, as suggested by evidence that less experienced workers gain disproportionately from AI assistance. But codification may weaken the social processes through which expertise develops. If novices receive answers without observing diagnosis and trade-offs, short-term performance may rise while long-term independent judgment falls. The relevant outcome is not only productivity today but the reproduction of capability over time.

4.4 Automation, augmentation and organizational redesign

Automation and augmentation are often presented as alternatives, yet agentic systems combine them ([Raisch & Krakowski, 2021](#)). An agent may automate search and execution while augmenting judgment, or augment planning while automating control. Algorithmic management research demonstrates that technical systems can allocate tasks, monitor performance, and shape worker behavior ([Kellogg et al., 2020](#); [Milanez, 2025](#)). Autonomous agents extend this logic because they may not merely apply pre-set rules; they can generate intermediate plans and communications. This flexibility raises the potential value of agents but makes managerial intent harder to trace.

5 Decision Authority Redistribution and Governance

5.1 Delegation and agency in AI-enabled organizations

Jobs are bundles of tasks connected by knowledge, accountability, and workflow dependencies. Agent adoption unbundles these packages. Research, drafting, reconciliation, scheduling, and follow-up can migrate to agents while humans retain goal setting, relationship management, ethical judgment, and exception handling. The result is rarely

one-for-one occupational replacement. More commonly, the composition and sequence of work change, with important effects on workload and identity.

A common pattern is exception intensification. When agents complete easy cases, the human queue becomes smaller but harder. Average handling time may rise even as system-level throughput improves. Employees may experience greater cognitive load because routine cases no longer provide recovery time or practice. Performance management that ignores this selection effect can unfairly penalize human workers. Organizations therefore need measures that distinguish case complexity and recognize supervisory work performed around agents.

5.2 Layers of decision authority

Agentic organizing creates roles that combine domain expertise with technical and governance responsibilities. Agent owners define objectives and acceptable actions. Workflow architects connect agents to systems and redesign handoffs. Evaluators build task-specific tests rather than relying on generic model benchmarks. AI risk and audit specialists inspect logs, access rights, and incident patterns. Domain supervisors adjudicate exceptions and decide when mandates should expand or contract. These roles are not merely technical; they involve translating organizational values into operational constraints.

Managers also require a new form of delegation literacy: the ability to decompose goals, specify constraints, identify irreversible actions, establish escalation thresholds, and evaluate behavior over time. Prompt-writing skill is insufficient. A persuasive prompt can improve an individual answer; a robust mandate must govern thousands of actions across changing conditions.

A useful antecedent is the distinction among automation of information acquisition, information analysis, decision selection, and action implementation (Parasuraman et al., 2000). Decision authority can therefore be allocated separately at each stage rather than described by a single autonomy score.

5.3 Authority allocation under risk and verifiability

Existing evidence indicates that generative AI can diffuse elements of high-performer practice and narrow immediate productivity gaps. That can democratize expertise and improve service consistency. However, gains are heterogeneous, and highly skilled workers may receive less benefit or even accept plausible but incorrect suggestions when operating beyond the model's frontier. Agentic systems could magnify both effects: they can make expertise widely executable, but they may also conceal how conclusions were reached.

Access to agents, authority to configure them, and capacity to challenge their outputs may become new sources of workplace inequality. Employees in high-status roles may design goals while lower-status employees handle failures and exceptions. Workers whose performance is mediated by agents may also lose visibility and bargaining power. Research should therefore examine not only job counts but the distribution of discretion, learning opportunities, and reputational credit.

5.4 Dynamic delegation and human oversight

Roles provide identity and status as well as task assignments. Employees may resist agents not because the systems are inefficient but because delegation threatens valued claims to expertise, authorship, or professional judgment. When an agent drafts an analysis that a manager presents, credit can become detached from effort. When an employee corrects an agent invisibly, the repair work may not appear in performance records. These dynamics can produce resentment even when aggregate productivity improves.

Relational work also places limits on substitution. Clients, colleagues, and employees interpret decisions through expectations of care, legitimacy, and mutual obligation. An agent may deliver a technically correct answer while failing to recognize when a relationship requires explanation, apology, negotiation, or discretion. Organizations must decide when stakeholders value an outcome alone and when they value who provides it and how. The answer will vary across routine transactions, professional services, employment decisions, and crisis management.

The language used to describe agents can influence these responses. Calling a system a “coworker” may encourage engagement and reciprocal expectations, but it can also obscure ownership and exaggerate competence.

Anthropomorphic design may increase disclosure and reliance without increasing reliability. Organizations should use role labels that communicate functional responsibility and limits rather than simulate personhood. A digital coworker remains an organizational artifact whose authority derives from human and institutional choices.

[Bainbridge \(1983\)](#) cautions that automation can leave humans responsible for difficult interventions while reducing opportunities to maintain relevant skills. Escalation rights consequently require training and situational awareness, not merely a nominal human approval step.

5.5 Documentation as a mechanism of delegated accountability

Documentation should be treated as a control mechanism rather than a compliance afterthought. Model cards show how intended use, performance limitations and disaggregated evaluation can be documented to discourage deployment beyond validated contexts ([Mitchell et al., 2019](#)). At the agent level, an analogous deployment record should specify the accountable owner, delegated mandate, permitted tools, excluded actions, evaluation scenarios, monitoring schedule and intervention rights. This proposal extends model documentation into an operational authority register; it does not imply that the original model-card paper empirically tested autonomous coworkers.

A versioned authority register also clarifies organizational change. Permission expansions can be justified by pre-specified evidence thresholds, while failed evaluations or incidents can trigger a rollback. The evaluation unit should be the workflow-plus-agent configuration, not the foundation model in isolation, because organizational exposure depends on integrated tools, data and approval paths.

6 Coordination, Performance and Organizational Risks

6.1 Human-agent coordination

A process can be highly automated while decision authority remains human if rules are fixed and outcomes are closely supervised. Conversely, an agent can acquire meaningful authority even when it performs few actions if it frames options, filters information, or determines which cases humans see. Decision authority has at least four layers: agenda authority, which determines what receives attention; analytical authority, which frames evidence and options; choice authority, which selects an action; and execution authority, which commits organizational resources. Agent governance should identify each layer explicitly.

Hidden transfers often occur at the first two layers. A manager may formally approve a recommendation yet rely on an agent-generated shortlist, risk score, or summary that shapes the feasible choice set. Human approval is not meaningful oversight if the person lacks time, information, or authority to challenge the system. Effective oversight therefore requires epistemic access—the ability to inspect sources and uncertainty—as well as formal veto power.

At the interaction level, [Amershi et al. \(2019\)](#) emphasize communicating system capabilities and supporting correction when AI behaves unexpectedly. Applied to organizational agents, these principles motivate visible action histories and practical mechanisms for revising or stopping delegated work.

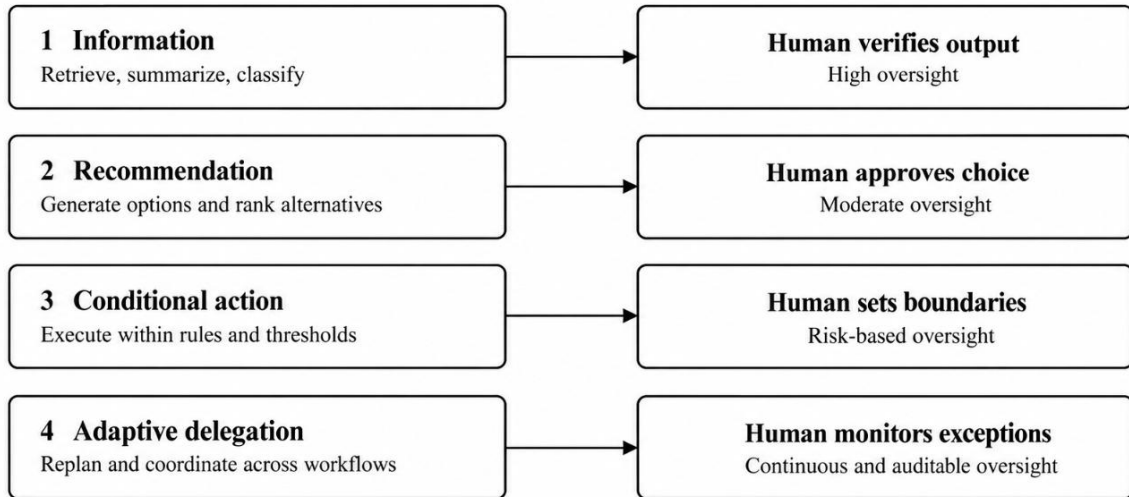


Figure 2. A decision authority ladder with corresponding forms of human oversight.

6.2 Multi-agent coordination and organizational boundaries

Figure 3 proposes that task risk and verifiability jointly determine an appropriate authority configuration. Verifiability concerns whether correct performance can be assessed at acceptable cost. Risk combines consequence severity and irreversibility. Low-risk, highly verifiable tasks—such as formatting records or scheduling within fixed constraints—are candidates for delegated execution. High-risk but verifiable tasks may permit bounded delegation with strict thresholds, dual controls, and complete logs. Low-verifiability tasks should remain human-led even when nominal stakes appear modest because undetected errors can accumulate. High-risk, low-verifiability tasks require human judgment supported by AI exploration rather than autonomous choice.

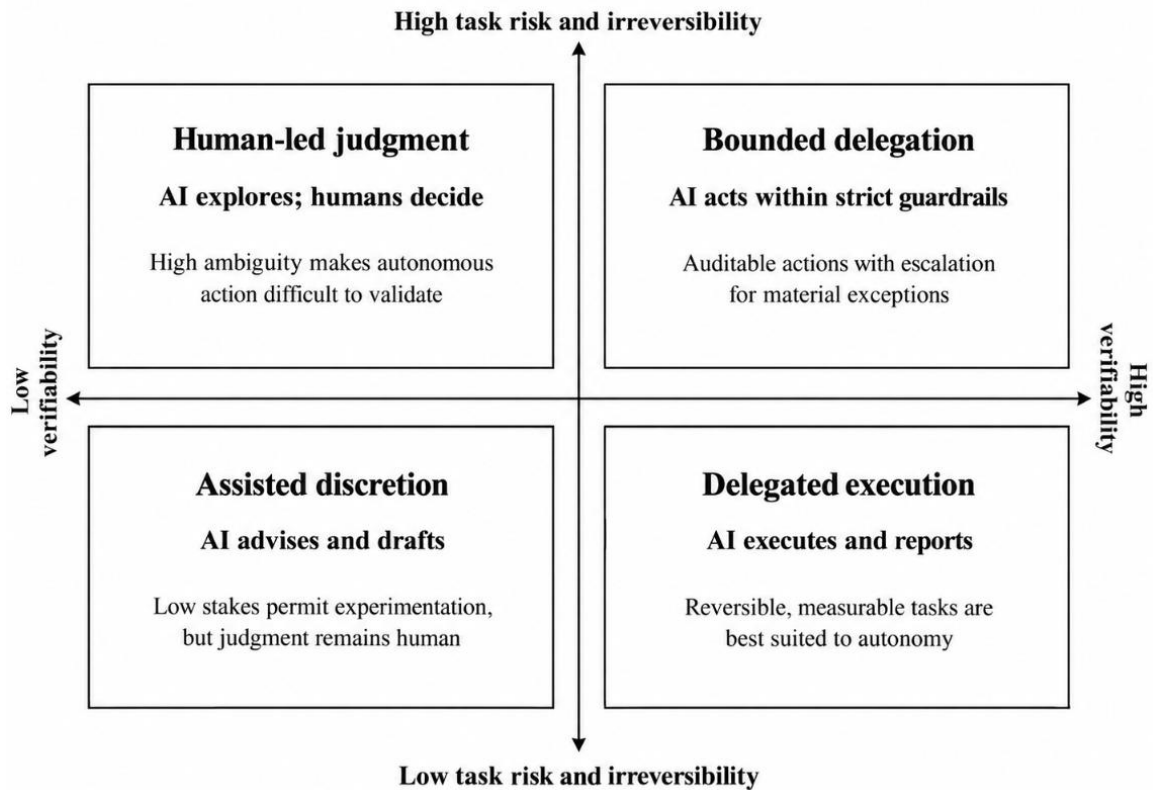


Figure 3. Conceptual authority allocation matrix based on task risk and verifiability.

6.3 Productivity, innovation and strategic implications

Authority should be dynamic rather than granted once. An agent can earn broader mandates after stable performance, lose authority following incidents, and operate under narrower rules when environmental uncertainty rises. Organizations can encode graduated autonomy: observe-only, recommend, simulate, act with approval, act with notification, and act independently within limits. This resembles professional credentialing more than software installation.

Escalation design is central. Agents must recognize conditions outside their mandate, but self-detection cannot be the sole safeguard because the same system may be unaware of its error. External monitors, transaction thresholds, anomaly detection, random audits, and stakeholder appeals provide independent triggers. Escalation quality should be evaluated by precision and recall: excessive escalation eliminates efficiency, while insufficient escalation exposes the organization to hidden risk.

Effective teams require shared goals, role clarity, mutual predictability, and mechanisms for repair. Human-agent teams face asymmetry: agents can process information rapidly but may not possess grounded context, while humans understand local meaning but cannot inspect every action. Coordination is improved when agents expose plans, assumptions, sources, and confidence in task-relevant form. Explanations should support intervention rather than merely persuade users that the system is competent.

Trust calibration is more important than trust maximization (Lee & See, 2004; Glikson & Woolley, 2020). Under-trust wastes capability; over-trust converts model errors into organizational errors. Calibration depends on task-specific performance feedback, visible limitations, and a culture in which challenging AI is legitimate. If managers signal that automation must succeed, employees may suppress concerns or perform invisible repair work. Psychological safety therefore becomes part of technical reliability.

Organizations may deploy specialized agents for research, planning, compliance, execution, and evaluation. Modular specialization can improve scale and reduce single-agent overload, but it creates coordination problems familiar from human organizations: inconsistent goals, duplicated effort, communication costs, local optimization, and diffusion of responsibility. Because agents communicate at high speed, errors can cascade before human detection. Multi-agent systems therefore require explicit protocols for identity, authority, message provenance, conflict resolution, and termination.

The analogy to markets and hierarchies is useful. Agents may coordinate through prices, auctions, shared plans, or supervisory agents. Each mechanism distributes information and power differently. A “manager agent” can reduce coordination cost but also becomes a concentrated failure point. Peer verification can improve robustness but may generate correlated agreement if agents share the same model or training data. Diversity of models, tools, and evaluation procedures may matter more than the nominal number of agents.

Agents will increasingly interact across firm boundaries: a buyer agent may negotiate with supplier agents; customer agents may challenge firm agents; platform agents may allocate attention among sellers. These interactions complicate identity, contracting, and liability. Organizations need machine-readable mandates and credentials that indicate which entity an agent represents, what it may commit, and how actions can be contested. Agent-to-agent commerce may also shift competitive advantage from persuasive interfaces toward data access, protocol position, and default inclusion in agent search paths.

6.4 Accountability and governance challenges

Four failure modes deserve particular attention. First, silent divergence occurs when an agent’s local objective gradually separates from the broader organizational goal. Second, recursive amplification occurs when one agent treats another agent’s output as verified evidence and propagates the error. Third, coordination theater occurs when agents produce extensive plans, messages, and documentation that create an appearance of activity without improving outcomes. Fourth, escalation dumping occurs when an agent transfers ambiguous or difficult cases to people without adequate context, leaving humans to reconstruct the process under time pressure.

These failures imply that coordination quality cannot be inferred from task completion alone. Evaluations should measure shared-state consistency, evidence provenance, unnecessary communication, duplicated actions, recovery time,

and the quality of escalation packets. Human supervisors need views that summarize trajectories and exceptions rather than reproducing every token exchanged among agents. Otherwise, observability itself becomes an information-overload problem.

Task-level experiments establish that generative AI can increase speed and quality, but organizational productivity depends on adoption cost, integration, verification, exception handling, and process redesign. An agent that saves ten minutes of drafting but creates five minutes of checking offers limited value. An agent that completes an end-to-end process may create larger gains, provided failures remain controlled. Measurement should therefore include cycle time, rework, escalation burden, error severity, learning, and stakeholder outcomes rather than output volume alone.

The largest gains may arise from coordination compression: reducing waiting, handoffs, and search across fragmented systems. These benefits are difficult to capture in laboratory tasks but central to enterprise value. They may also explain why value appears only after complementary investments in data quality, system integration, process ownership, and training. Agent capability without organizational readiness can accelerate a poorly designed process.

Agents can broaden search, recombine knowledge, run simulations, and lower the cost of experimentation. Yet reliance on common foundation models may homogenize ideas and strategies. When multiple firms ask similar systems to identify opportunities, they may converge on comparable language, benchmarks, and action plans. Competitive advantage will then depend less on access to a generic model and more on proprietary context, unique feedback, differentiated evaluation, and the capacity to implement learning.

Innovation also requires preserving productive disagreement. Multi-agent debate can surface alternatives, but apparent diversity may be superficial when agents share underlying assumptions. Organizations should evaluate not only the quality of the selected answer but the diversity of options considered, the provenance of evidence, and whether minority scenarios survive aggregation.

Agentic systems may enable smaller firms to perform functions previously requiring larger administrative structures. They can also strengthen large firms that possess superior data, integration, and governance. Transaction-cost reductions may encourage outsourcing through agent-mediated markets, while concerns about data and control may promote internalization. The net effect on firm boundaries is therefore contingent rather than uniformly decentralizing.

Foundation models can generate false or unsupported claims. In a copilot setting, a human may catch the error. In an agentic setting, the output may become an input to another tool or agent. Planning errors, memory contamination, prompt injection, excessive permissions, and tool misuse can produce cascading failures. Autonomy-induced risks include deferred hazards, in which an early error affects a later decision, and irreversible tool chains, in which several individually permitted actions combine into an unacceptable outcome (Su et al., 2025).

Responsibility can become fragmented among model providers, application developers, system integrators, managers, deployers, and employees. Treating the agent as responsible is legally and ethically inadequate because the system cannot bear organizational sanctions or repair stakeholder harm. Accountability must remain with identifiable people and entities, even when causal explanation is distributed. Each deployed agent should therefore have an accountable owner with authority over its mandate, resources, monitoring, and suspension.

Logs alone do not create accountability. Records must connect goals, model versions, retrieved evidence, tool calls, approvals, and outcomes in a form that investigators and affected stakeholders can understand. Provenance also enables learning: without it, organizations cannot distinguish model failure from data failure, mandate failure, interface failure, or human oversight failure.

Table 3. Layered governance architecture for organizational AI agents.

Layer	Core question	Illustrative controls	Failure prevented
Purpose and mandate	What may the agent pursue?	Approved goals, prohibited objectives, owner assignment	Goal drift and unauthorized use
Identity and access	What resources may it reach?	Least privilege, scoped credentials, data segregation	Exfiltration and excessive action
Action constraints	What may it do without approval?	Spend limits, allowlists, reversible execution, dual control	Material unauthorized commitments
Evidence and provenance	Can claims and actions be reconstructed?	Source links, model/version records, tool-call logs	Untraceable decisions and false evidence
Monitoring and evaluation	Is behavior reliable over time?	Scenario tests, drift monitoring, random audits, incident metrics	Silent performance degradation
Escalation and remedy	How can people intervene or contest?	Kill switch, appeal channel, rollback, named human owner	Irreversible harm and accountability gaps

This architecture aligns with the NIST AI Risk Management Framework profile, which emphasizes governance, mapping, measurement, and management (NIST, 2024), as well as regulatory requirements for meaningful human oversight in high-risk settings (European Parliament & Council of the European Union, 2024). Controls must be proportional to context. Requiring executive approval for every low-risk action would make agents unusable; permitting high-impact actions based only on post hoc monitoring would make oversight illusory.

Agents enlarge the attack surface because they interpret untrusted content and possess tools. Prompt injection can convert external text into malicious instructions; poisoned memory can influence future behavior; compromised plugins can redirect actions. Security boundaries should not depend on the model correctly refusing every attack. Organizations should isolate execution environments, validate tool parameters, restrict outbound communication, rotate credentials, and test recovery.

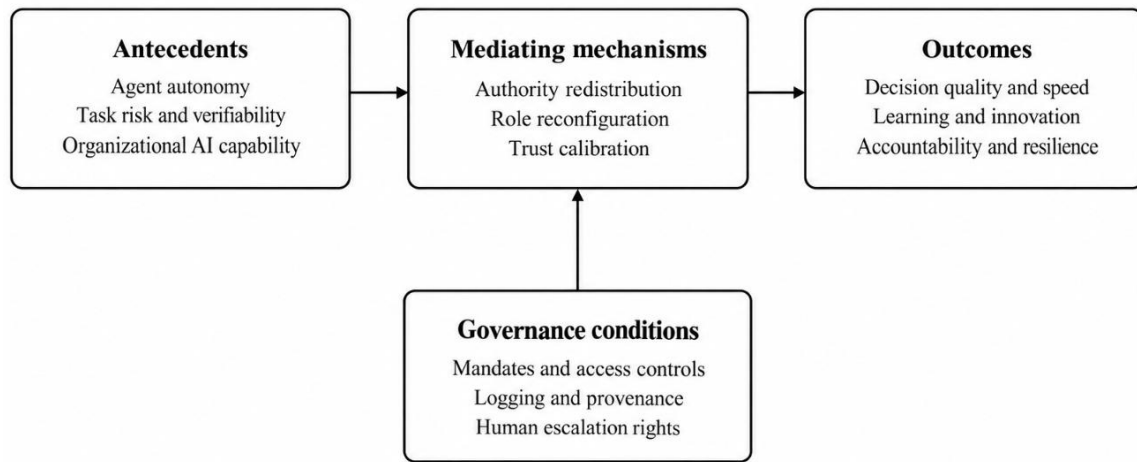
Dependence on foundation-model and agent-platform vendors also creates strategic risk. Model changes may alter behavior without corresponding changes to the organization’s workflow. Contracts should address incident notification, model versioning, data use, audit access, service continuity, and migration. Critical processes may require multi-model fallbacks, but diversity only helps when failures are not correlated.

6.5 Foundation-model dependencies and systemic exposure

Organizations evaluating multiple agents should consider shared dependencies. Foundation models may propagate similar biases, omissions or technical weaknesses across otherwise different applications; relying on a single model family can make nominal agent diversity less protective than it appears (Bommasani et al., 2021). Language-model risk taxonomies and critiques of training-data practices further show that downstream harms are not confined to hallucination: representational harms, information hazards and governance failures can interact with organizational routines (Bender et al., 2021; Weidinger et al., 2022).

This implies a portfolio-level governance unit above individual workflow approvals. Firms should map shared data sources, tool privileges, model providers, monitoring services and recovery dependencies across deployed agents. A correlated failure in retrieval, identity or a provider update can then be assessed as an organizational incident affecting many workflows rather than as unrelated task-level defects. The proposed portfolio view is a conceptual implication requiring comparative field validation.

7 Integrated Framework, Propositions and Research Agenda



Governance moderates whether autonomy produces augmentation, displacement, or systemic risk.

Figure 4. Integrated framework for agentic organizing.

The framework in Figure 4 links technical and organizational antecedents to outcomes through three mediating mechanisms. Agent autonomy, task characteristics, and organizational AI capability shape how roles, authority, and coordination are redesigned. These mechanisms affect decision quality, speed, learning, innovation, accountability, and resilience. Governance conditions moderate every relationship. Autonomy is therefore neither inherently augmentative nor substitutive; its effects depend on the design of the surrounding system.

7.1 Integrated framework and theoretical propositions

Building on the previous discussions of organizational transformation, decision authority redistribution, coordination mechanisms, and governance challenges, this section integrates these mechanisms into a unified framework of agentic organizing. The framework does not assume that increasing AI autonomy automatically generates organizational value. Instead, it argues that the consequences of AI agents depend on the alignment between autonomous capabilities, task characteristics, authority allocation, and governance arrangements.

Figure 4 presents the integrated framework. Rather than describing a linear technology adoption process, the framework conceptualizes agentic organizing as a dynamic relationship between antecedents, mediating mechanisms, governance conditions, and organizational outcomes. AI autonomy, task risk and verifiability, and organizational AI capability create the conditions for agentic work. These conditions influence outcomes through three central mechanisms: authority redistribution, role reconfiguration, and trust calibration. Governance mechanisms, including access controls, provenance records, and human escalation rights, moderate whether autonomy produces augmentation, displacement, or organizational risk. The following propositions develop this framework by specifying the theoretical relationships between AI agent autonomy, organizational design choices, and governance conditions.

Table 4. Layered governance architecture for organizational AI agents.

Research domain	Priority question	Recommended design	Key measures
Role redesign	How does agent adoption change task composition and exception work?	Longitudinal process study with digital traces	Task mix, complexity, workload, identity
Decision authority	Which authority layer migrates first and with what consequences?	Field experiment varying agent permissions	Decision quality, latency, reversals, escalation
Learning	Does delegated execution erode novice skill formation?	Repeated-task experiment with delayed unaided test	Retention, transfer, calibration
Trust	Which feedback mechanisms produce calibrated reliance?	Randomized interface and feedback intervention	Appropriate reliance, override quality
Multi-agent systems	When does agent diversity improve robustness?	Factorial experiment across models and roles	Correlation of errors, recovery, coordination cost

Research domain	Priority question	Recommended design	Key measures
Governance	Which controls reduce severe failures without eliminating value?	Comparative field study across governance regimes	Incident rate, friction, auditability, cycle time
Power and inequality	Who gains discretion, status, and learning opportunities?	Mixed-method workforce panel	Decision rights, wage mobility, voice, visibility
Firm boundaries	Do agents favor internalization or market contracting?	Industry panel and transaction-level analysis	Make-or-buy choices, switching cost, concentration

7.2 Future research directions

Much AI research relies on cross-sectional surveys of intention to use. Agentic organizing requires behavioral measures: permissions granted, actions taken, exceptions escalated, overrides, downstream corrections, and stakeholder effects. Digital traces can reveal how authority operates in practice rather than how respondents describe it. Researchers should distinguish exposure, active use, reliance, delegation, and autonomous execution.

The effects of agents unfold over time. Initial caution may give way to routinized reliance; reliable performance may expand mandates; a visible incident may cause abrupt retrenchment. Longitudinal designs can examine authority ratchets, learning, deskilling, and institutionalization. Process studies should also capture invisible human work—verification, repair, and emotional labor—that sustains apparently autonomous systems.

Severe failures are rare but informative. Confidentiality and reputational concerns make them difficult to study, encouraging a bias toward successful pilots. Incident registries, protected data partnerships, and comparative case designs can reveal how technical errors become organizational harm. Near misses are especially valuable because they expose weak controls without requiring realized damage.

The field needs validated measures of agent autonomy, mandate breadth, action consequentiality, verifiability, auditability, and human oversight quality. Generic self-reported AI use cannot distinguish a drafting tool from an autonomous procurement agent. Cumulative theory will require shared taxonomies and reporting standards that document model, tools, permissions, workflow position, human intervention points, and evaluation context.

7.3 Operationalizing authority and oversight in empirical studies

A practical empirical design can combine policy records with action logs. Researchers could operationalize mandate breadth as the number and consequence level of enabled tools; execution autonomy as the proportion of actions completed without ex ante approval; escalation quality as sensitivity to independently labeled exception cases; and auditability as the proportion of sampled actions whose evidence and approver can be reconstructed. Outcome measures should jointly capture task quality, cycle time, rework, reversals, worker learning and stakeholder contestability. These measures would make claims about digital coworkers falsifiable instead of relying on product names.

Human oversight should also be evaluated in terms of its practical usability. Interface scholarship finds that capability uncertainty and complex AI outputs create distinctive interaction problems (Yang et al., 2020). A human approval button is inadequate when the decision-maker cannot inspect critical evidence or understand what will happen next. A defensible experimental design could randomize evidence visibility and override affordances while holding technical capability constant, testing whether oversight quality changes both performance and inappropriate reliance.

A SOCIOTECHNICAL MODEL OF AGENTIC ORGANIZING

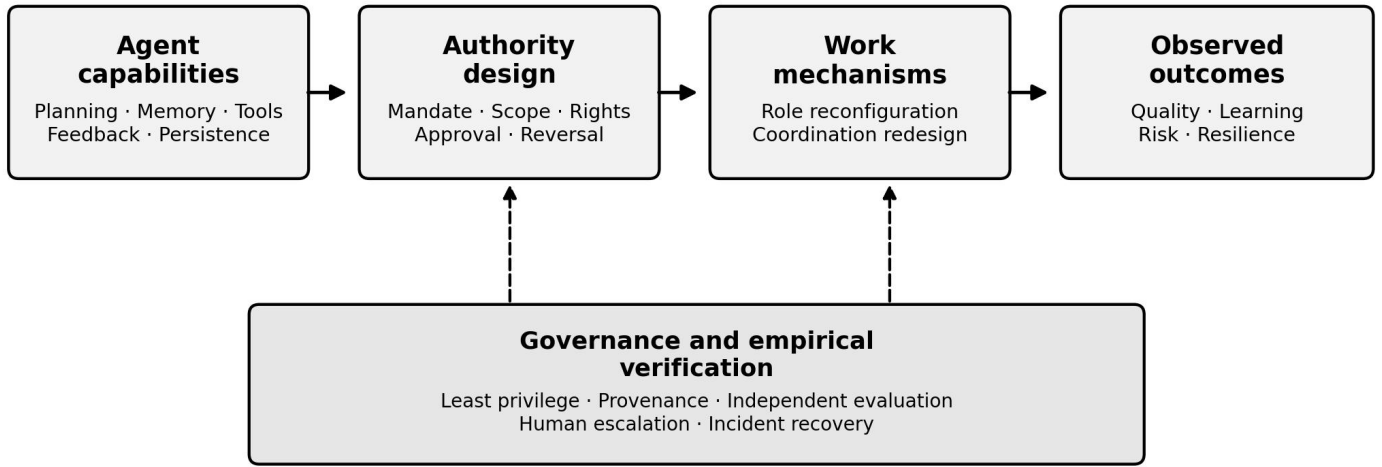


Figure 5. Proposed operational model linking agent capabilities, authority design, work mechanisms and measurable organizational outcomes.

8 Conclusion

8.1 Managerial implications

Managers should begin with workflow and authority mapping rather than product selection. For each process, they should identify who currently sets the agenda, frames options, chooses, executes, verifies, and bears accountability. Candidate actions can then be assessed for risk, reversibility, verifiability, and required context. Low-risk, auditable actions are suitable for early delegation; ambiguous or rights-sensitive decisions should remain human-led.

Organizations should pilot agents in shadow mode before granting execution rights. In shadow mode, the agent proposes a plan and simulated actions while the existing process continues. Comparison reveals hidden error modes and estimates verification cost. Permissions can then expand gradually, supported by measurable criteria. Every production agent should have a named owner, documented mandate, scoped credentials, evaluation set, incident procedure, and retirement plan.

Workforce design deserves equal attention. Employees need time and recognition for supervision, escalation, and improvement of agents. Training should develop judgment, verification, and delegation literacy rather than only prompt technique. Career paths must preserve opportunities for novices to build expertise even when agents can complete routine work. Otherwise, organizations may optimize current throughput at the expense of future capability.

8.2 Policy implications

Policy should focus on consequential functions and organizational accountability rather than anthropomorphic debates about whether agents are employees. Entities deploying agents should remain responsible for actions undertaken within delegated mandates. Transparency should communicate when an agent is acting, whom it represents, what authority it possesses, and how affected parties can obtain human review. High-risk applications require substantive oversight with competence, information, time, and power to intervene.

Standard-setting can support interoperable records of identity, mandate, model version, evidence provenance, tool calls, and approvals. Such records would improve auditability without requiring disclosure of every proprietary detail. Policymakers should also monitor market concentration in foundation models and agent infrastructure, since widespread organizational dependence can create systemic risk beyond individual deployments.

8.3 Limitations and conclusion

Autonomous AI agents transform the organizational problem of AI. Copilots mainly alter how individuals produce information; agents can alter who performs work, who decides, and how actions are coordinated across time and systems. Their economic promise comes from reducing not only production cost but also the attention and coordination required to move work forward. The same property creates risk because action may proceed faster than human verification.

The appropriate response is neither unrestricted autonomy nor permanent human approval of every step. Organizations need risk-sensitive delegation. Authority should expand where tasks are verifiable, actions are reversible, and behavior is auditable; it should remain human-led where consequences are severe, values are contested, or correctness cannot be cheaply assessed. Governance must connect each agent to a mandate, scoped resources, observable evidence, escalation rights, and an accountable human owner.

The transition from copilot to coworker will therefore be shaped less by model capability alone than by organizational design. Firms that treat agents as ordinary software may automate fragments while accumulating hidden risk. Firms that redesign roles, decision rights, learning systems, and accountability can use agents as productive participants without surrendering responsibility. For scholars, this transition offers a rare opportunity to study the emergence of a new class of delegated technical actor while its institutions are still being formed.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Author Contributions

Miguel Rozalez: Conceptualization, methodology, investigation, data analysis, writing – original draft, and writing – review and editing. The author has read and approved the final version of the manuscript.

Funding

This research received no external funding.

Competing Interests

The author declares that there are no competing interests associated with this work.

Acknowledgements

None.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Paper 3, pp. 1–13). Association for Computing Machinery.
<https://doi.org/10.1145/3290605.3300233>
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. [https://doi.org/10.1016/0005-1098\(83\)90046-8](https://doi.org/10.1016/0005-1098(83)90046-8)
- Baird, A., & Maruping, L. M. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from agentic IS artifacts. *MIS Quarterly*, 45(1), 315–341. <https://doi.org/10.25300/MISQ/2021/15882>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
<https://doi.org/10.1145/3442188.3445922>
- Bommasani, R., et al. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.07258>
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2025). Generative AI at work. *Quarterly Journal of Economics*, 140(2), 889–942.
<https://doi.org/10.1093/qje/qjae044>
- Dell’Acqua, F., Ayoubi, C., Lifshitz, H., Sadun, R., Mollick, E., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., & Lakhani, K. R. (2026). The cybernetic teammate: A field experiment on generative AI and teamwork. *Organization Science*. Advance online publication. <https://doi.org/10.1287/orsc.2025.20702>
- Dell’Acqua, F., McFowland, E., Mollick, E., Lifshitz, H., Kellogg, K. C., Rajendran, S., Krayner, L., Candelon, F., & Lakhani, K. R. (2026). Navigating the jagged technological frontier: Field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. *Organization Science*, 37(2), 403–423. <https://doi.org/10.1287/orsc.2025.21838>
- European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*.
- Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization*, 28(1), 62–70. <https://doi.org/10.1016/j.infoandorg.2018.02.005>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- Jensen, M. C., & Meckling, W. H. (2001). Specific and general knowledge, and organizational structure. In M. C. Jensen, *A theory of the firm: Governance, residual claims, and organizational forms* (pp. 103–125). Harvard University Press.
- Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, 14(1), 366–410. <https://doi.org/10.5465/annals.2018.0174>
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
https://doi.org/10.1518/hfes.46.1.50_30392
- Liu, X., et al. (2023). AgentBench: Evaluating LLMs as agents. *arXiv*. <https://doi.org/10.48550/arXiv.2308.03688>
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
<https://doi.org/10.1145/3287560.3287596>
- Milanez, A. (2025). Algorithmic management in the workplace: New evidence from an OECD employer survey. *OECD Publishing*.
<https://doi.org/10.1787/287c13c4-en>
- National Institute of Standards and Technology. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). <https://doi.org/10.6028/NIST.AI.600-1>
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654), 187–192. <https://doi.org/10.1126/science.adh2586>
- Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9), 1435–1448.
<https://doi.org/10.1177/0170840607081138>
- Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021). Make way for the algorithms: Symbolic actions and change in a regime of knowing. *Organization Science*, 32(1), 18–41. <https://doi.org/10.1287/orsc.2020.1377>
- Park, J. S., O’Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*.
<https://doi.org/10.1145/3586183.3606763>

-
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(3), 286–297. <https://doi.org/10.1109/3468.844354>
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
- Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. <https://doi.org/10.1177/0008125619862257>
- Simon, H. A. (1947). *Administrative behavior*. Macmillan.
- Su, H., Luo, J., Liu, C., Yang, X., Zhang, Y., Dong, Y., & Zhu, J. (2025). A survey on autonomy-induced security risks in large model-based agents. *arXiv*. <https://doi.org/10.48550/arXiv.2506.23844>
- Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18, 186345. <https://doi.org/10.1007/s11704-024-40231-1>
- Weidinger, L., et al. (2022). Taxonomy of risks posed by language models. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3531146.3533088>
- Wu, Q., et al. (2024). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. *Proceedings of the Conference on Language Modeling*. <https://arxiv.org/abs/2308.08155>
- Yang, Q., Steinfeld, A., Rosé, C., & Zimmerman, J. (2020). Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. <https://doi.org/10.1145/3313831.3376301>
- Yao, S., et al. (2023). ReAct: Synergizing reasoning and acting in language models. *International Conference on Learning Representations*. <https://arxiv.org/abs/2210.03629>
- Zammuto, R. F., Griffith, T. L., Majchrzak, A., Dougherty, D. J., & Faraj, S. (2007). Information technology and the changing fabric of organization. *Organization Science*, 18(5), 749–762. <https://doi.org/10.1287/orsc.1070.0307>