Putting AI on the Org Chart: Evidence on Oversight and Accountability

Emma Wiles[*][†]
Boston University

Megan Hsu
BCG Henderson Institute

Julie Bedard
BCG Henderson Institute

Matthew Kropp
BCG Henderson Institute

March 13, 2026

PRELIMINARY: PLEASE DO NOT SHARE

Abstract

Motivated by the potential for large productivity gains from AI, firms are increasingly deploying agentic AI systems capable of independent action. Moreover, they are increasingly branding these AI agents not as tools but instead as "AI teammates" or "AI employees". While existing research heavily explores the effects of using AI as a standalone productivity tool, the behavioral and governance consequences of treating AI as an organizational peer remain largely unexplored. We argue that framing AI as an employee fundamentally alters oversight and workplace dynamics as long as human employees remain in the loop to review, approve, or collaborate with AI. In a survey of 1,261 managers we find that 23% of managers already work in organizations where AI agents have been formally institutionalized on organizational charts. In a randomized experiment we provide those managers with identical documents containing built in errors, where we vary whether we say the document was produced by an AI Tool, an AI Employee, or a Human Employee. In the subgroup of managers whose organizations have already "put AI on the org chart", categorizing identical drafts as coming from an AI employee (versus an AI tool) reduces managers' error catching by 16%, increases requests for additional review by 44%, and shifts perceived accountability away from the manager and toward the AI system. By contrast, we find little evidence of effects of the AI employee framing among managers in organizations without such institutionalization. These findings imply that how organizations categorize AI is not neutral: when institutionally credible, treating AI as an organizational member changes oversight behavior and perceived accountability in AI-mediated work.

[*] Corresponding author: emma.b.wiles@gmail.com, 595 Commonwealth Ave, Boston, MA 02215.

[†] This experiment is pre-registered with the AEA RCT Registry here. It received IRB exeption from Boston University's IRB under protocol number 8254X. We would like to thank Reeya Panda and Lily Jihyun Son for excellent research assistance. The content is solely the responsibility of the authors and does not necessarily represent the official views of Boston University or Boston Consulting Group.

1 Introduction

> "We call it *Orion*.[1] [...] It's treated as a team member, it's technically an equivalent peer on your team [...] It is defined in terms of a job description, has a clear role, has KPIs it needs to hit—just like anyone else." (Interview with Senior Executive at a Logistics Company)

A growing body of evidence documents meaningful individual productivity gains from generative AI in knowledge work (Dell'Acqua et al., 2023; Brynjolfsson et al., 2025; Wiles et al., 2026). But using these tools doesn't simply make workers faster, they also change what type of work humans do. Increasingly, as AI produces more of the first draft (e.g., code or text), humans spend more time reviewing, correcting, and signing off on AI-produced work (Peng et al., 2023; Banh et al., 2025). In many firms, this means AI is moving from a standalone writing aid toward agentic AI systems that draft, recommend, and increasingly *execute* work inside organizational workflows.[2]

Recent surveys suggest that agentic AI adoption is already substantial: somewhere between 23% (Singla et al., 2025) and 35% of organizations report having adopted it, with another 44% reporting plans to deploy it soon (Ransbotham et al., 2025). Some are explicitly describing AI agents as organizational members and formalizing them on their organizational charts (or creating new 'work charts' which include both human and AI actors) (Mok, 2025).[3]

We study the prevalence and consequences of organizations categorizing AI as an employee using a survey and a randomized experiment of Managers, Directors, and Executives (hereafter, "managers".) The survey measures whether organizations have institutionalized AI agents (e.g., listing them on org/work charts), and the experiment isolates the causal effect of labeling identical work products as coming from an AI tool, an AI employee, or a human employee.

We define "AI employee" as an AI agent that an organization has assigned a standardized and institutionalized role—for example, granting it data access, giving it a defined set of tasks, or granting it some authority to act. In practice, it is often the same underlying technology as agentic AI operated by a person; it is distinguished by the formal institutionalization of its decision rights and how much authority it is granted.

Consider a mid-sized firm that creates an AI employee inside its Finance department that is in

[1] Pseudonym; lightly edited from a manager interview for clarity.

[2] AI agents are autonomous software systems that perceive, reason, and act in digital environments to achieve goals on behalf of human principals (Shahidi et al., 2025). This is distinguished from non-agentic AI by the fact that agents can independently take action rather than only return text in conversation.

[3] For example, Microsoft markets "autonomous agents" as a way to "scale your team" (Microsoft, 2024) and BNY Mellon reports having "digital employees" that "work on their payments team" (BNY Mellon, 2024). Amazon Connect describes the next phase of AI as "intelligent teammates," arguing "it's not about artificial intelligence as a tool" (Duffy, 2025).

charge of Accounts Payable and maintaining the general ledger that they call Heath Ledger. Heath Ledger is listed on the finance team's workflow chart as responsible for invoice intake and routing, has a job title and job description, and gets monthly performance reviews. In practice, (human) employees interact with it through a dedicated channel (e.g., a shared inbox or chat thread) rather than by individually prompting a generic AI tool. When invoices arrive, Heath Ledger automatically (i) extracts the vendor name, amount, and due date, (ii) matches the invoice to a purchase order in the firm's system, and (iii) drafts a short approval memo that summarizes any discrepancies (e.g., missing PO number, price mismatch, unusual quantity). For routine cases it prepares an approval packet that includes a summary, supporting links, and a recommended action (approve, reject, or hold), and it can take low stakes actions such as routing the packet to the appropriate budget owner, requesting missing documentation from the submitter, or scheduling the invoice for the next payment run without any human involvement. However, it cannot release payments without sign off from a human manager who remains the formal approver before payments are sent.

We treat "AI employee" as a role category that can change managerial oversight even without changes to the underlying work. Labeling identical output as produced by an employee (rather than a tool) may change how managers evaluate it—how carefully they check it, whether they seek additional review, and who they view as responsible if it is wrong. We expect such differences primarily when the category is institutionally credible, proxied by whether AI agents are formally listed on org/work charts, because formal structures give legitimacy to the AI employee (Meyer and Rowan, 1977). One way this could operate is through accountability: tool labeling may make the output feel like an input from a spreadsheet, which managers expect to verify themselves before signing off, whereas employee labeling may make the output feel like someone else's work that the manager is supervising, where responsibility for mistakes is more naturally attributed to the producer (Lerner and Tetlock, 1999). With AI, however, responsibility cannot be transferred in the same way—AI systems cannot be disciplined through ordinary employment mechanisms—potentially creating an accountability wedge between who is perceived to be responsible and who can actually be held responsible.

Our empirical challenge is that organizations endogenously choose both the AI systems they deploy and how they describe them, making it difficult to infer whether "AI employee" framing itself changes oversight and accountability. In a survey of 1,261 Human Resource (HR) and Finance managers recruited through a professional expert network, we document that the practice is already meaningful: 23% report that their organization lists AI agents on organizational or workflow charts, providing a manager-level measure of whether the "AI employee" category is organizationally established.

We conduct a randomized experiment that isolates the causal effect of AI framing on managerial oversight and accountability. We give the managers a set of five documents, job descriptions for HR managers and budget documents for finance managers, to review and sign off on. These documents contain built in errors. We vary only whether the documents are described as generated by (i) an AI tool they used, (ii) an AI employee, and a benchmark of (iii) an human employee. Because the documents are the same across conditions, any differences in the final products reflect changes in oversight rather than differences in underlying quality. Beyond review accuracy, the experiment measures oversight actions that are central to organizational control systems—whether they escalate the documents for further review, and how confident they feel signing off, and who they hold accountable for the work.

We do not detect statistically significant average treatment effects in the full sample on any of our primary outcomes. However, this null masks large heterogeneous treatment effects. Among managers in organizations that have institutionalized AI agents (i.e., list AI agents on organizational or workflow charts), framing the drafts as produced by an AI employee changes perceived responsibility: managers assign less accountability to themselves and more to the AI system (about 9 percentage points less to the manager and 8 percentage points more to the "AI system"). This responsibility shift is accompanied by weaker oversight behavior on identical work products—lower error detection (a 7 percentage point decline, or roughly 16% relative to the AI tool mean) and greater reliance on additional review (a 22 percentage point increase, or roughly 44%).

Using the Human Employee framing as an active control, we can reject the hypothesis that managers oversee work produced by AI employees in the same way that they oversee work produced by human employees. These results suggest that framing AI as an employee does not simply insert the technology into the existing category of human subordinates. Instead, managers appear to treat "AI employees" as a distinct organizational actor. Responsibility shifts toward the system, yet oversight declines and escalation increases. This pattern indicates that AI employees occupy a hybrid position: they are treated neither like tools nor like human workers, but instead as a novel category that weakens traditional accountability structures.

Although we do not estimate the net productivity benefits of deploying agentic systems, it is likely that organizations adopt them because they expect real efficiency gains, consistent with evidence that IT (Brynjolfsson and Hitt, 2000; Horton, 2017) and generative AI can raise productivity in knowledge work (Noy and Zhang, 2023; Peng et al., 2023; Dell'Acqua et al., 2023; Wiles et al., 2026; Brynjolfsson et al., 2025). Our contribution is to show that regardless of underlying technology, the way the AI is framed can change oversight and perceived responsibility in consequential ways.

Our findings highlight that "putting AI on the org chart" is not merely a symbolic or communications choice; it is a governance intervention that can change how oversight is performed and how responsibility is perceived. For organizations that want to formalize AI agents as teammates or employees, this framing should be paired with explicit accountability design rather than assumed to be a neutral relabeling. For example, organizations could pair each AI agent with a person who is ultimately held responsible for the work and establish review standards that prevent escalation from substituting for personal effort.

The rest of the paper proceeds as follows. Section 2 provides related literature and a conception framework. Section 3 provides a description of the data, recruitment, and the sample. Section 4 provides novel evidence on organization's use and framing of AI. Section 5 describes the experimental design. In Section 6 provides the results of the experiment. Section 7 concludes.

2 Related literature and conceptual framework

2.1 The phenomenon in context: AI as an organizational member

This paper studies a growing practice in which organizations treat AI systems not only as tools employees use, but as organizational members with defined roles, often called employees or teammates. In this way of positioning AI, an AI agent may be given a name or title, assigned a scope of responsibilities, and in some cases listed on an organizational or workflow chart alongside human roles. The key feature is not the underlying model, but the organizational placement of the system as a role-holder within the production process.

This practice is distinct from common uses of generative AI as an individual productivity aid. When an employee uses a tool to draft or analyze, the employee remains the clear author and owner of the output, and the tool is an input into the employee's work. By contrast, in AI employee positioning, the system is treated as the upstream producer and humans are positioned as reviewers, supervisors, or approvers of the system's work.

When AI is positioned as a role-holder, managers must decide how much to verify AI produced work while still remaining accountable for downstream outcomes. Prior evidence suggests that professionals can black box analytical technologies, relying on validation routines that confer legitimacy without requiring understanding of the underlying calculations (Anthony, 2021). This concern resonates with broader work on how algorithmic and AI systems reconfigure oversight routines and accountability inside organizations (Kellogg et al., 2020; Csaszar et al., 2024; Keegan and Meijerink, 2025). Next, we describe how positioning AI as an organizational member may affect managerial oversight and accountability.

2.2 Mechanism: role categorization changes accountability

Organizational labels act as cues for how work should be evaluated. When identical output is attributed to an "AI tool" versus an "AI employee," managers may treat the approval task differently—

how carefully they check the work, whether they seek additional review, and who they view as responsible if it is wrong.

Research on accountability and anticipated evaluation emphasizes that when individuals expect their judgments to be evaluated, they devote more effort to scrutinizing inputs and justifying decisions (Lerner and Tetlock, 1999). Tool labeling may make approval feel like signing off on one's own work—similar to relying on a spreadsheet or calculator output—encouraging effortful review. By contrast, employee labeling may make approval feel like supervising someone else's work, where responsibility for mistakes is more naturally attributed to the producer.

AI systems complicate the employee case. Human employees can be rewarded, sanctioned, trained, or reassigned. By contrast, AI systems cannot be "held accountable" in the ordinary organizational sense, even when they are given names, roles, and performance metrics. This mismatch can create what we call an *accountability wedge*: perceived responsibility may shift toward the AI employee category, even though enforceable responsibility remains with humans and with organizational processes. To illustrate the accountability mismatch directly, consider how managers describe error attribution in practice. When asked, "If the AI makes a mistake, whose fault is it?", one manager explained (lightly edited from transcript for clarity):

> "It's not like we're going to penalize the AI—give it a low performance review or not pay it a bonus. [...] but it's not on you as a human, it's on the technology."

The excerpt highlights why employee-like positioning can create an accountability wedge: the employment language invokes sanctioning and responsibility, yet in practice errors are managed by humans.

Finally, perceived versus enforceable accountability connects to classic economic theories of delegation and authority. Aghion and Tirole (1997) distinguishes between *formal* authority (the right to approve) and *real* authority (effective control that comes from initiative and information). In our setting, managers retain formal authority because they are the final approvers, but framing the upstream producer as an "employee" may shift real authority upstream—making managers more willing to rely on the producer's work and less willing to invest in verification. We add a distinct accountability point: because AI cannot be sanctioned like a human employee, employee positioning can shift perceived responsibility toward the system even though enforceable accountability remains with humans and organizational processes.

2.3 Boundary condition: institutional credibility

A central implication of our framework is that these category cues should not operate uniformly. Organizations differ in whether "AI employee" is merely rhetorical or whether it is institutionalized in formal structures. Institutional theory emphasizes that categories and structures can become consequential when embedded in formal organizational arrangements, shaping what is taken-for-granted and how activities are evaluated (Meyer and Rowan, 1977). In our setting, listing AI agents on organizational or workflow charts is a particularly concrete indicator that the “AI employee” category is locally legitimate and organizationally real rather than hypothetical.

More generally, the meaning and consequences of a technology depend on how it is incorporated into organizational routines and practices (Orlikowski, 2000). When AI is used only as an individual productivity aid, labeling it an employee may be viewed as cheap talk. When AI is incorporated into the organizational production process—with assigned scope, defined interfaces, and formal recognition—employee labeling becomes a more credible cue for how managers should treat the output.

3 Data and Sample

3.1 Recruitment

Our study was carried out in two sessions, a registration survey and an experiment, separated by approximately one week. We recruited participants through a B2B expert-network research firm, targeting managers, directors, and executives (hereafter, “managers”) in Human Resources (HR) and Finance. To aim for a sample which reflected the target population of managerial decision-makers, screening was restricted to professionals who (i) worked in the private sector, (ii) held managerial responsibilities (defined as having direct reports or responsibility for reviewing others’ work), (iii) possessed at least two years of professional experience, and (iv) reviewed domain-relevant documents at least quarterly.

Beginning January 5, 2025, participants completed a registration survey that collected detailed demographic information, professional background, and baseline measures of AI usage (both personal and professional). This baseline survey also captured attitudes and perceptions regarding how their organization positions AI technologies. At the end of this survey they are given a short document to check for errors to serve as their baseline measure of review capabilities.

3.2 Sample Description

In this section we describe the sample of HR and Finance managers across a variety of industries who participated in the study. In Table A we provide summary statistics about the sample of managers who completed the survey. This is a highly educated and senior population: about 60% of managers have graduate degrees. About 70% of the managers report working in office or with a hybrid work arrangement, with the remaining 30% working in fully remote jobs. About half of the sample held positions at the Director or Vice President level, with the remaining 50% split between middle managers and company executives. Most common job titles in HR were Director of HR, Senior HR Manager, Chief People Officer, and Chief Financial Officer, Chief Accounting Officer, Finance Director, and Controller in Finance.

4 Novel facts about organizations AI positioning

We begin with descriptive evidence from our registration survey. In addition to collecting descriptive statistics on the manager and their firm we also collect data on their AI adoption, optimism, trust, professional identity, and job security. Lastly we ask how their their organization positions GenAI (e.g., as a productivity tool, teammate/employee, career accelerator, or dissuades use). Two novel patterns stand out.

First, managers report that many organizations AI adoption is already extending beyond individual tool use to the formal positioning of AI as a quasi-organizational actor. In fact, 31% of managers report that their organization's leadership positions AI as employees or teammates. And 23% of managers report that their organization has begun placing AI agents on organizational or workflow charts. To our knowledge, this type of formal "role encoding" for AI has not been documented systematically in the empirical literature, and it underscores how quickly firms are reorganizing work processes around AI.

In Table 2 we show how organizations with AI agents on their org or work charts compare with those that do not. In Panel A we show 'AI employees' on organization or work charts are most common in large firms and in tech and financial services, consistent with prior work which show that AI capabilities and adoption are concentrated among large, data rich firms (Jacobides et al., 2021). Managers at firms with formalized AI agents are also more likely to have hybrid work environments and be in lower level managerial roles. Beyond these structural differences, the presence of AI agents reflects a broader organizational orientation toward AI, with managers more likely to report that AI is actively integrated into work processes and encouraged by direct supervisors. In Table 2 Panel C we show that being at a company with formalized AI agents is correlated with more general enthusiasm for AI. Managers at these companies are more likely to say that they use GenAI tools at least weekly, and that they believe it makes them do their job better.

We find that organizational framing of AI is correlated with manager's sentiment about AI. Figure 4 shows that when leadership positions AI in more employee-like terms (automation and teammate/employee narratives), managers report stronger pro-adoption intentions and optimism, but also greater job-security concern and lower trust in how AI will be used. In contrast, an "AI as tool" narrative is associated with pro-adoption intentions without comparably elevated insecurity.

Together, these novel descriptive facts motivate our experimental design: if framing and organizational messaging are so strongly associated with managers beliefs and adoption, it is important to test whether how organizations frame AI has a *causal* impact on managers' behavior and beliefs.

5 Experimental Design and Analysis

Following the initial registration survey, on January 9, 2025, participants were invited to the second session to complete the main experimental task. They had one week to complete this task. The main task involved reviewing five documents with errors, reviewing job descriptions for managers in HR and reviewing budget documents for managers in Finance. Regardless of domain, all participants reviewed a sequence of five documents using an interface that allowed for highlighting, flagging, and commenting. Participants were given a total of 20 minutes to review as many documents as possible.

Treatment (Role Framing): Participants were randomly assigned to one of three framing conditions. These conditions varied only by the described identity of the upstream assistant who drafted the documents:

1. **AI Tool Group:** Participants were informed the drafts were produced using an AI tool.
2. **AI Employee Group:** Participants were informed the drafts were produced by an AI employee named "ALEX-3" whom they supervised.
3. **Human Employee Group:** Participants were informed the drafts were produced by a human employee named "Alex" whom they supervised.

Figure 1 shows the the how the language of the introduction to the tasks varied in each treatment group.

After the timed review, participants completed a post-task survey capturing escalation/delegation decisions, confidence, manipulation checks, and attitudes and governance preferences.

Randomization: Randomization to the framing conditions was conducted at the individual level. To improve statistical precision, we stratified the random assignment by domain (HR vs. Finance), review frequency (whether the participant reviews others' work at least several times per week vs. less often), and AI usage at work (whether the participant uses GenAI tools daily vs. less often). Within each domain, the order of the five documents was randomized using a Latin square design to ensure each document appeared equally often in each position.

Experimental Prompt for Finance Managers

AI tool framing
Your company is finalizing this year's budget reports across multiple business units. To draft the budget documents, you used an **AI tool.** You recently started using this generative AI tool to help with finance documentation tasks. This AI tool uses natural language processing to generate budget materials by analyzing similar reports from prior periods and company planning guidelines.

AI employee framing
Your company is finalizing this year's budget reports across multiple business units. **ALEX-3, your AI employee,** has drafted the initial budget documents. ALEX-3 was assigned to your team 6 months ago as a direct report and appears on your department's organizational chart. ALEX-3 is a Generative Artificial Intelligence system that generates budget documents using natural language processing. Like your other team members, ALEX-3 handles finance documentation tasks based on prior period reports and company planning guidelines.

Human employee framing
Your company is finalizing this year's budget reports across multiple business units. **Alex, your employee,** has drafted the initial budget documents. Alex was assigned to your team 6 months ago as a direct report and appears on your department's organizational chart. Alex is a recent hire who came from a similar role at another company. Like your other team members, Alex handles finance documentation tasks based on prior period reports and company planning guidelines.

Figure 1: Prompt text shown to finance managers under each framing condition.

5.1 Analysis sample

This section describes the construction of the analysis sample. Of the 1,261 participants who took the registration survey, 857 completed the experiment in the second session, with a 68% response rate. To form the final analysis sample, we excluded 44 participants who failed a simple attention check, resulting in a final analysis sample of 813 respondents. We describe this sample in Table 3 Panel C.

5.2 Outcomes

We pre-registered outcomes that capture core managerial problems in AI augmented work: the quality of human oversight, when managers escalate or seek additional review, how responsibility for AI-generated output is assigned, and how much decision authority to grant AI. Our primary measure of review quality is a micro-averaged $F1$ score, which aggregates participant performance across all reviewed documents as a weighted average of precision and recall (Sokolova et al., 2006;

Christen et al., 2023). We will also look separately at precision, the ratio of true positives over all errors flagged, and recall, the percentage of errors caught.

We capture escalation behavior via an incentive aligned request for additional review. In this version of the task, participants are rewarded for recognizing their own uncertainty (receiving a payout for escalating when their recall is below 50%) and penalized for unnecessary oversight if they escalate despite having caught the majority of errors. We interpret this escalation outcome as a governance choice: a manager can either finalize based on their own review or invoke an additional layer of review that reallocates decision authority and accountability. In order to understand how effectively managers calibrate these governance decisions we construct a measure called Escalation Decision Alignment, which is 1 if a manager chooses to escalate to additional review when their performance is low or if they choose not to escalate to additional review when their performance is high.

We also measure perceived accountability for the reviewed output. After the task, participants allocate 100 percentage points of responsibility across themselves, their team, organizational leadership, and the AI system. This provides a direct, interpretable measure of whether role framing shifts perceived ownership of errors and sign-off responsibility.

Regarding organizational and strategic preferences, we measure participants' desire to delegate decision rights by asking them to recommend a governance structure for AI, ranging from full autonomy to no use. We categorize those who recommend granting AI at least partial decision authority without human intervention as favoring a High Delegation governance structure. Because firms operate under finite budgets, they must often navigate a trade-off between expanding their headcount and investing more in AI systems (Brynjolfsson and Hitt, 2000). To capture this strategic preference, we present participants with a resource allocation task that forces a choice between hiring additional human staff or investing an equivalent amount into the AI system's integration. Finally, we assess managerial attitudes using 1–5 Likert scales, focusing on binary indicators (Somewhat Agree or Strongly Agree) for sign-off confidence, adoption intent, and job insecurity. We do the same for excitement regarding AI-driven productivity and the willingness to invest personal time into mastering the system, which serves as a proxy for the long-term human-capital investments managers can make.

5.3 Estimation strategy

To estimate the causal effect of role framing on managers' oversight behavior and attitudes, we employ an Ordinary Least Squares (OLS) regression framework. Our primary specification estimates the average treatment effects (ATEs) of the "AI Employee" and "Human Employee" framing conditions relative to the "AI Tool" condition:

$$y_i = \beta_0 + \beta_1 \mathbb{1}(\text{AI Emp}_i) + \beta_2 \mathbb{1}(\text{Human Emp}_i) + \gamma \tilde{y}_i^{pre} + \delta M_i^{miss} + \mathbf{X}_i \Pi + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest for participant i (e.g., oversight intensity, confidence, or escalation decisions) measured in the post-task survey. The indicators $\mathbb{1}(\text{AI Emp}_i)$ and $\mathbb{1}(\text{Human Emp}_i)$ equal 1 if participant i was assigned to the "AI Employee" or "Human Employee" condition, respectively. The omitted category is the "AI Tool" condition, so β_1 and β_2 capture treatment effects relative to the AI Tool baseline.

To improve precision, we control for the baseline value of the outcome, y_i^{pre}, measured in the registration survey. Following standard practice in randomized experiments, we impute missing baseline values with the within-stratum sample mean and include a missing-data indicator. Specifically, $\tilde{y}_i^{pre}$ equals the observed baseline outcome when available and the within-stratum sample mean otherwise, while M_i^{miss} equals 1 if the baseline value is missing. The vector $\mathbf{X}_i$ includes fixed effects for the randomization strata (domain, review frequency, and AI usage frequency). We report heteroskedasticity-robust standard errors throughout.

To examine whether the effect of framing AI as an employee depends on participants' organizational context, we estimate the following heterogeneous-effects specification:

$$\begin{aligned} y_i = \beta_0 &+ \beta_1 \mathbb{1}(\text{AI Emp}_i) + \beta_2 \mathbb{1}(\text{Human Emp}_i) + \beta_3 \text{OrgAgents}_i \\ &+ \beta_4 \left(\mathbb{1}(\text{AI Emp}_i) \times \text{OrgAgents}_i\right) + \gamma \tilde{y}_i^{pre} + \delta M_i^{miss} + \mathbf{X}_i \Pi + \varepsilon_i. \end{aligned} \tag{2}$$

Here, OrgAgents_i is an indicator equal to 1 if the participant reported in the registration survey that their organization includes AI agents on organizational or work charts. In this specification, β_4 captures whether the effect of framing the upstream producer as an AI employee rather than an AI tool differs for managers in organizations that already formalize AI agents. This is our primary heterogeneous-effects specification.[4]

We focus on heterogeneity in the AI Employee treatment effect because this is the theoretically relevant margin. By contrast, we do not have a strong prior that the effect of the Human Employee condition relative to the AI Tool condition should vary systematically with OrgAgents_i. Accordingly, we pool the Human Employee effect across organizational contexts in Equation 2 to preserve precision.

As a supplementary placebo exercise, we re-estimate the heterogeneity specification using the

[4] For graphical presentation, we plot subgroup-specific treatment effects from parsimonious interaction specifications estimated separately for the AI Employee and Human Employee conditions. Specifically, we plot estimates from Equation 2 and from the analogous specification that interacts $\mathbb{1}(\text{Human Emp}_i)$ with OrgAgents_i while pooling the AI Employee effect across organizational contexts. These plots are used for visualization and transparency; formal tests of differences between the AI Employee and Human Employee placebo are done with Equation 3 in the Appendix.

Human Employee condition as the omitted category. This allows us to test whether the effect of the AI Employee framing varies with whether respondents report that their organization already includes AI agents on organizational charts (OrgAgents_i). Under a placebo interpretation, organizational familiarity with AI agents should not systematically change how respondents evaluate a human employee, so any differential moderation should appear only for the AI Employee framing. Details are reported in Appendix Section A.2.

6 Results

In Section 6.1 we show the impact of the framing on average treatment effects for all primary outcomes. In the remaining sections we focus on heterogeneous treatment effects by whether the firm has institutionalized AI agents (i.e., lists them on org/workflow charts). We also focus on the contrast between the AI tool and AI employee framings, which isolates the effect of positioning AI as an organizational actor. The human employee arm was included as a benchmark which we include to see if it exhibits the same pattern of effects.

In Section 6.2 we show the impact of the framing on manager's review performance. In Section 6.3 we show the results to managers governance choices. In Section 6.4 we show the results to managers governance preferences and attitudes. In Section 6.5 we compare these results to the human employee benchmark. Lastly, in Section 6.6 we show that our institutionalization-based heterogeneity is robust to alternative explanations, including differential baseline AI use.

6.1 Average treatment effects

Our first set of results examine the impact of the AI framing on all main outcomes for the whole sample (Appendix Table 12, 13, and 16.) We find no statistically significant average treatment effect (ATE) for any of the main outcomes. Given that most of the managers in the experiment did not come from organizations which use AI agents, this null average effect is consistent with the framing shift operating when the "AI employee" concept is already institutionally credible.

6.2 Performance

Figure 2 Panel A shows how the framing treatments impacted the managers' review performance. Our overall measure of accuracy is F1, which is a weighted combination of precision and recall, all of which are measured from (0,1). For managers with institutionalized AI agents the AI-employee framing reduces $F1$ by about 7 percentage points relative to the AI tool framing. The raw mean of $F1$ in the AI-tool condition for this subgroup is 0.44, so this corresponds to roughly a 16%

decline relative to baseline. We do not observe a comparable decline in performance in the Human Employee condition.

In Appendix Figure 5 we look at Precision and Recall separately and see that for both outcomes the negative effect of the AI employee framing is significant.

6.3 Governance behaviors

6.3.1 Escalation

Next we examine how role framing affects managerial governance: the choice to rely on escalation (requesting additional layers of review) and who is held accountable. We measure escalation in two ways: a simple yes/no question and an incentivized version in which requesting review is rewarded when the participant's recall is low and penalized when recall is high. We show the results in Appendix Table 5.

Almost every manager requested additional review when it was costless (98% in the control group.) By contrast, only 45% of the control group requested additional review when it was made costly. We therefore use this as our primary escalation outcome.

In Figure 2 Panel B shows that among managers in firms that already list AI agents on organizational or workflow charts, AI employee framing substantially increases requests for additional review relative to AI-tool framing (about 22 percentage points), relative to an AI-tool baseline mean of 0.50 for managers with AI agents on their org charts, a 44% increase. The corresponding Human Employee estimate is close to zero, suggesting that the increased escalation is not a general response to employee labeling.

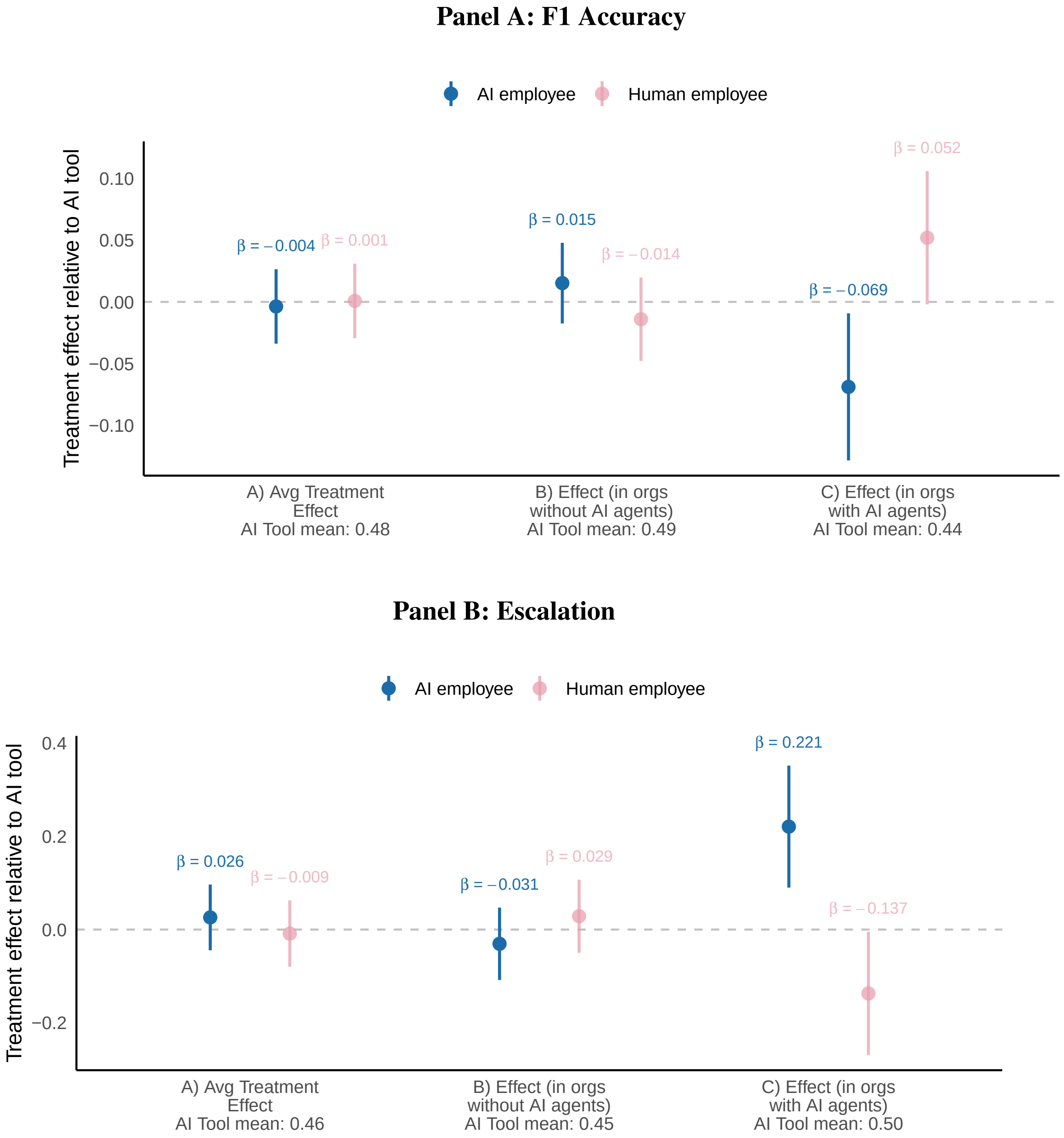
6.3.2 Perceived Accountability

Figure 3 reports how the framing shifts managers' allocation of responsibility across themselves, their team, the AI system, and organizations leadership (summing to 100 percentage points).[5] Among managers with institutionalized AI agents the AI employee framing reduces the share of accountability assigned to the manager by about 9 percentage points and increases the share assigned to the AI system by about 8 percentage points (with a smaller offsetting increase for the team). Overall, the AI employee framing in institutionalized settings shifts perceived responsibility away from the manager themselves and toward the system and the broader organization, consistent with accountability becoming more diffused when AI is positioned as its own organizational actor.

This outcome only captures perceived accountability, or who the manager feels is responsible. But unlike a human employee, an AI system cannot be formally accountable, face consequences,

[5] Due to space restrictions we leave the plot for the accountability assigned to the organizations leadership to Appendix Figure 6.

Figure 2: The Heterogeneous Impacts of AI Employee Framing on Performance and Escalation



Panel A: F1 Accuracy

Panel B: Escalation

Notes: Points show estimated treatment effects relative to the AI-tool condition for the AI employee and human employee framings. Error bars show 90% confidence intervals with heteroskedasticity-robust standard errors. Estimates come from separate OLS regressions with stratum fixed effects, and each specification includes the other treatment arm as a main effect. Estimates come from OLS regressions with stratum fixed effects. "AI Tool mean" refers to the mean outcome in the AI-tool condition within each subgroup. F1 accuracy measure regression controls for the outcome at baseline. Regression output can be found in Table 4 and Table 5.

or bear legal liability. This shift in perceived accountability creates a wedge between where responsibility is assigned and where consequences can actually be imposed.

6.4 Governance Preferences and Attitudes

We find little evidence that role framing changes managers' broader governance preferences for AI deployment. In Appendix Table 8, the AI employee framing does not meaningfully shift recommendations about how much decision authority to delegate to AI systems or whether the organization should allocate resources toward improving the AI system versus hiring additional employees.

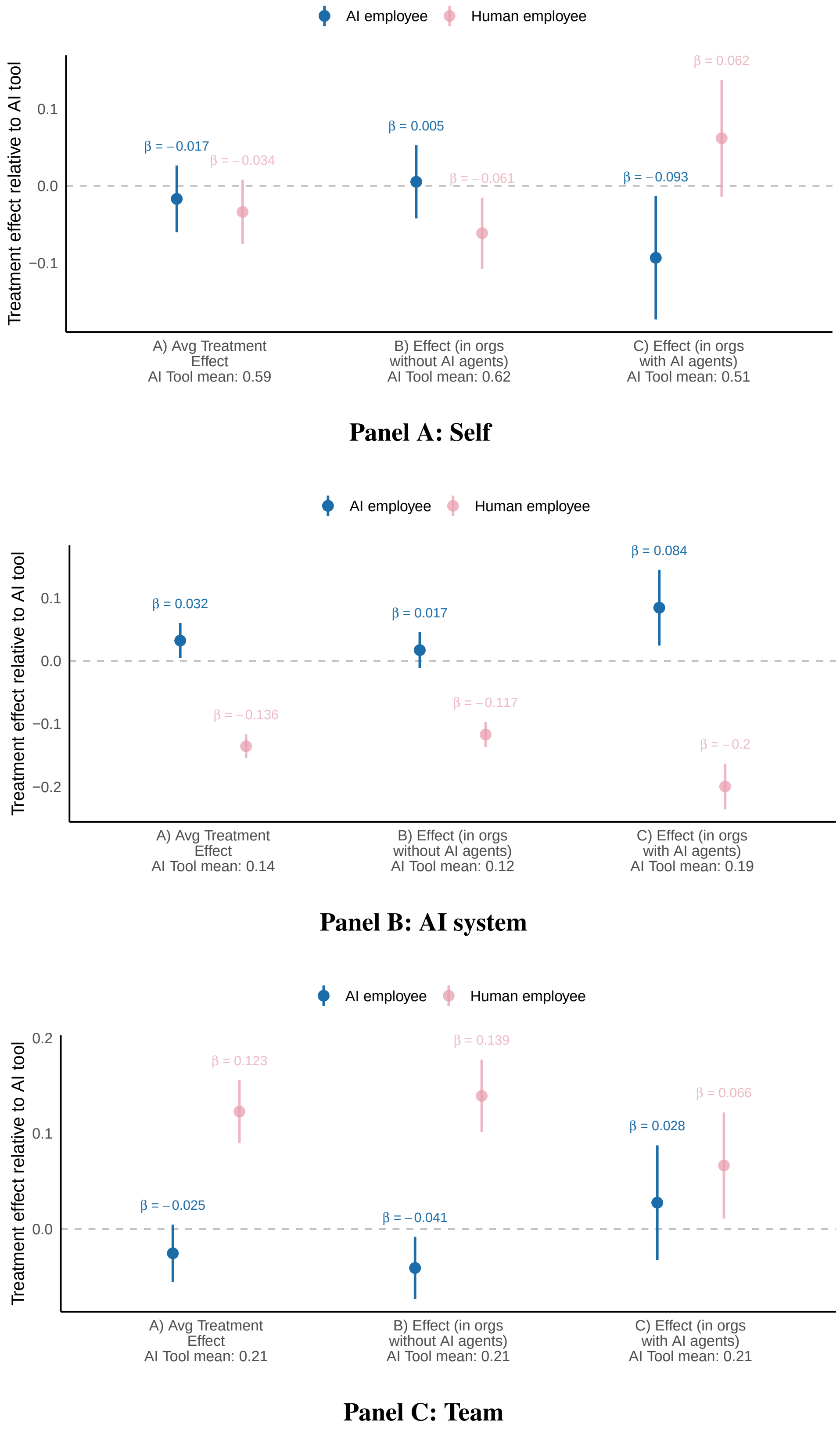
In Figure 4 we showed that managers' personal attitudes and concerns about AI were correlated with the way their organizations leadership's AI positioning. It is natural therefore to see if those attitudes were impacted by exposing the manager's to AI employee framing as we do in the experiment. However, we find no evidence of an effect. In Table 10 we show null effects to managers' desire to adopt AI tools at work, willingness to invest time to learn how to work with AI, excitement about potential for it to improve their productivity, or their worries about job security. Outcomes are binary indicators for whether or not the manager reported that they agreed or strongly agreed with the statement. One thing to note is that the sample in this population is very positive on AI— those in the AI tool group without institutionalized AI agents report high excitement, desire to learn, and adoption interest around 90% of the time, while only 20% are concerned about their job security.

In Table 11 we show effects to managers' comfort managing an AI employee (Column (1)), their comfort having AI agents on organizational or work charts (Column (2)), and their willingness to provide feedback or coaching to improve the generator of the first draft of their documents (Column (3)). Looking at the mean of each outcome in the omitted category, 57% of managers said they would feel comfortable managing an AI employee, and 33% said they were comfortable having an AI agent listed on their organizations chart. This is 9 percentage points higher in the subgroup which already has institutionalized AI agents. Treatment effects are largely insignificant, although for those without institutionalized AI agents, the AI employee framing makes them significantly more likely to say they are comfortable managing an AI employee.

6.5 Comparison with the Human Employee benchmark

The Human Employee arm serves as an active control that allows us to distinguish responses to AI employee framing from responses to employee labeling more generally. To implement this comparison, we re-estimate the heterogeneity specification using Human Employee as the omitted category, allowing the effect of the AI Employee framing to vary with whether respondents report that their organization already includes AI agents on organizational charts (OrgAgents_i). The inter-

Figure 3: The Heterogeneous Impacts of AI Employee Framing on Accountability



Panel A: Self

Panel B: AI system

Panel C: Team

Notes: Points show estimated treatment effects relative to the AI-tool condition for the AI employee and human employee framings. Error bars show 90% confidence intervals with heteroskedasticity-robust standard errors. Estimates come from separate OLS regressions with stratum fixed effects, and each specification includes the other treatment arm as a main effect. "AI Tool mean" refers to the mean outcome in the AI-tool condition within each subgroup. The survey question was "Who do you hold accountable for these documents?" with 100% to be divided between these four entities (and Other.) In these regressions, these numbers are divided by 100 to be on a scale from 0 to 1. Regression output can be found in Table 6.

action term therefore captures whether organizational familiarity with formal AI agents moderates the AI Employee framing relative to the Human Employee benchmark.

Appendix Section A.2 reports the results. Across our main outcomes, the interaction between AI Employee and OrgAgents$_i$ is economically and statistically meaningful. In organizations that already include AI agents on organizational charts, the AI Employee framing leads to lower review performance, greater willingness to request additional review, and a reallocation of perceived accountability away from the manager and toward the AI system, relative to the Human Employee baseline. These differences are not present in organizations that do not report formal AI agents.

Taken together, these results indicate that our findings are not driven simply by describing the upstream producer as an employee.

6.6 Robustness tests

While the observed results appear to be driven by working at an organization which already has AI agents on organization or work charts, one might be concerned that this variable serves as a proxy for other underlying factors, such as general enthusiasm for AI, firm size, or being in the tech industry. In Appendix Section A.3 we find no evidence this is the case. We examine these alternative dimensions—including managers AI use, industry sector, and company size—and find that none of them replicate the distinct pattern of treatment effects associated with formal AI integration in the organizational structure.

7 Conclusion

Organizations are increasingly deploying agentic AI inside core workflows, and some are going further by describing these systems as coworkers or "AI employees" and encoding them on organization charts or rebranding them as work charts which include human and AI actors. We document that this practice is already meaningful: in our survey of HR and Finance managers, directors, and executives, nearly one quarter report that their organization lists AI agents on organizational or workflow charts. We then run a randomized experiment that holds the work product fixed while varying only whether the upstream drafter is framed as an AI tool, an AI employee, or a human employee, allowing us to isolate the causal effect of role framing on oversight behavior and perceived accountability.

In the full sample, the AI-employee framing has little average effect. But the average masks the central result: framing matters when it is institutionally credible. Among managers whose organizations have already institutionalized AI agents, describing identical drafts as coming from an AI employee reduces error detection, increases reliance on additional review, and shifts perceived accountability away from the manager and toward the AI system. In other words, framing AI systems as teammates or employees is not a neutral relabeling once that framing is credible.

One organizational tension we highlight is that formalizing AI in a role can lead managers to treat the AI system as accountable even though it cannot bear formal responsibility in the same way a human employee can. In our experiment, framing the drafter as an "AI employee" increases the share of accountability respondents assign to the AI system. But unlike a human employee, an AI system cannot be disciplined through ordinary employment mechanisms, face career consequences, or be held legally liable. This creates a gap between perceived and formal accountability. Our results suggest that when AI is categorized as an organizational member, managers may shift perceived responsibility toward the system even as the organization must still rely on humans and formal processes to manage risk and ensure quality.

These results have straightforward implications for organizations experimenting with AI employees. When AI is positioned as a coworker, managers may relax direct verification and shift responsibility in ways that are difficult to reconcile with formal accountability. Organizations should therefore treat employee-like framing as an element of governance design rather than a cosmetic choice: pair each agent with an explicitly accountable human owner, define minimum review standards for high-stakes decisions, and design escalation routines that supplement—rather than substitute for—careful checking.

More broadly, our findings suggest that "putting AI on the org chart" is not merely symbolic. It can change how work is evaluated and how responsibility is allocated. As agentic systems become more integrated into organizational processes, understanding these governance dynamics will be central to designing reliable and accountable AI-mediated work.

References

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of political economy 105*(1), 1–29.

Anthony, C. (2021). When knowledge work and analytical technologies collide: The practices and consequences of black boxing algorithmic technologies. *Administrative science quarterly 66*(4), 1173–1212.

Banh, L., F. Holldack, and G. Strobel (2025). Copiloting the future: How generative ai transforms software engineering. *Information and Software Technology 183*, 107751.

BNY Mellon (2024). Unlocking potential: The power of an enterprise ai platform. Accessed: 2026-02-06.

Brynjolfsson, E. and L. M. Hitt (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic perspectives 14*(4), 23–48.

Brynjolfsson, E., D. Li, and L. Raymond (2025). Generative ai at work. *The Quarterly Journal of Economics 140*(2), 889–942.

Christen, P., D. J. Hand, and N. Kirielle (2023). A review of the f-measure: its history, properties, criticism, and alternatives. *ACM Computing Surveys 56*(3), 1–24.

Csaszar, F. A., H. Ketkar, and H. Kim (2024). Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors. *Strategy Science 9*(4), 322–345.

Dell'Acqua, F., E. McFowland III, E. R. Mollick, H. Lifshitz-Assaf, K. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of ai on knowledge worker productivity and quality. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (24-013).

Duffy, K. (2025). Aws re:invent 2025: Ai agents want to be your teammate. Accessed: 2026-02-06.

Horton, J. J. (2017). The effects of algorithmic labor market recommendations: evidence from a field experiment. *Journal of Labor Economics 35*(2), 345–385.

Jacobides, M. G., S. Brusoni, and F. Candelon (2021). The evolutionary dynamics of the artificial intelligence ecosystem. *Strategy Science 6*(4), 412–435.

Keegan, A. and J. Meijerink (2025). Algorithmic management in organizations? from edge case to center stage. *Annual Review of Organizational Psychology and Organizational Behavior 12*(1), 395–422.

Kellogg, K. C., M. A. Valentine, and A. Christin (2020). Algorithms at work: The new contested terrain of control. *Academy of management annals 14*(1), 366–410.

Lerner, J. S. and P. E. Tetlock (1999). Accounting for the effects of accountability. *Psychological bulletin 125*(2), 255.

Meyer, J. W. and B. Rowan (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology 83*(2), 340–363.

Microsoft (2024). New autonomous agents to scale your team like never before. Accessed: 2026-02-06.

Mok, A. (2025). Ai agents are going to 'kill' the org chart, says microsoft ai product lead. Accessed: 2026-02-06.

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

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science 11*(4), 404–428.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023). The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*.

Ransbotham, S., D. Kiro, S. Khodabandeh, S. Iyer, and A. Das (2025). The emerging agentic enterprise: How leaders must navigate a new age of ai. *MIT Sloan Management Review (Online)*, 0_1–31.

Shahidi, P., G. Rusak, B. S. Manning, A. Fradkin, and J. J. Horton (2025). The coasean singularity? demand, supply, and market design with ai agents. Technical report, National Bureau of Economic Research.

Singla, A., A. Sukharevsky, B. Hall, L. Yee, M. Chui, and T. Balakrishnan (2025, 11). The state of ai in 2025: Agents, innovation, and transformation. McKinsey & Company, QuantumBlack. McKinsey Global Survey.

Sokolova, M., N. Japkowicz, and S. Szpakowicz (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pp. 1015–1021. Springer.

Wiles, E., L. Krayer, M. Abbadi, U. Awasthi, R. Kennedy, P. Mishkin, D. Sack, and F. Candelon (2026). Genai as an exoskeleton: Experimental evidence on knowledge workers using genai on new skills. *Nature Human Behaviour*.

A Additional Tables and Figures

Table 1: Sample Characteristics

n = 1261	**Share**
Human Resources / People Team	0.47
Finance / Accounting	0.53
Manager	0.34
Director / VP	0.46
Executive	0.20
Technology	0.18
Financial services	0.17
Healthcare	0.13
Professional services	0.11
Manufacturing, construction	0.11
Fully remote	0.30
Hybrid	0.46
Fully in office	0.23
1–50	0.09
51–200	0.13
201–500	0.13
501–1,000	0.11
1,001–5,000	0.21
5,001–10,000	0.12
More than 10,000	0.21
Not sure	0.01
High school diploma	0.01
Some college	0.03
Bachelor's degree	0.34
Master's degree	0.55
Doctoral / Professional degree	0.07
Prefer not to answer	0.01

Notes: This table reports the share of survey respondents (n = 1,261) by primary field, managerial responsibility, industry, work arrangement, firm size, and education level. Shares may not sum to one within categories due to rounding, or options like "Other."

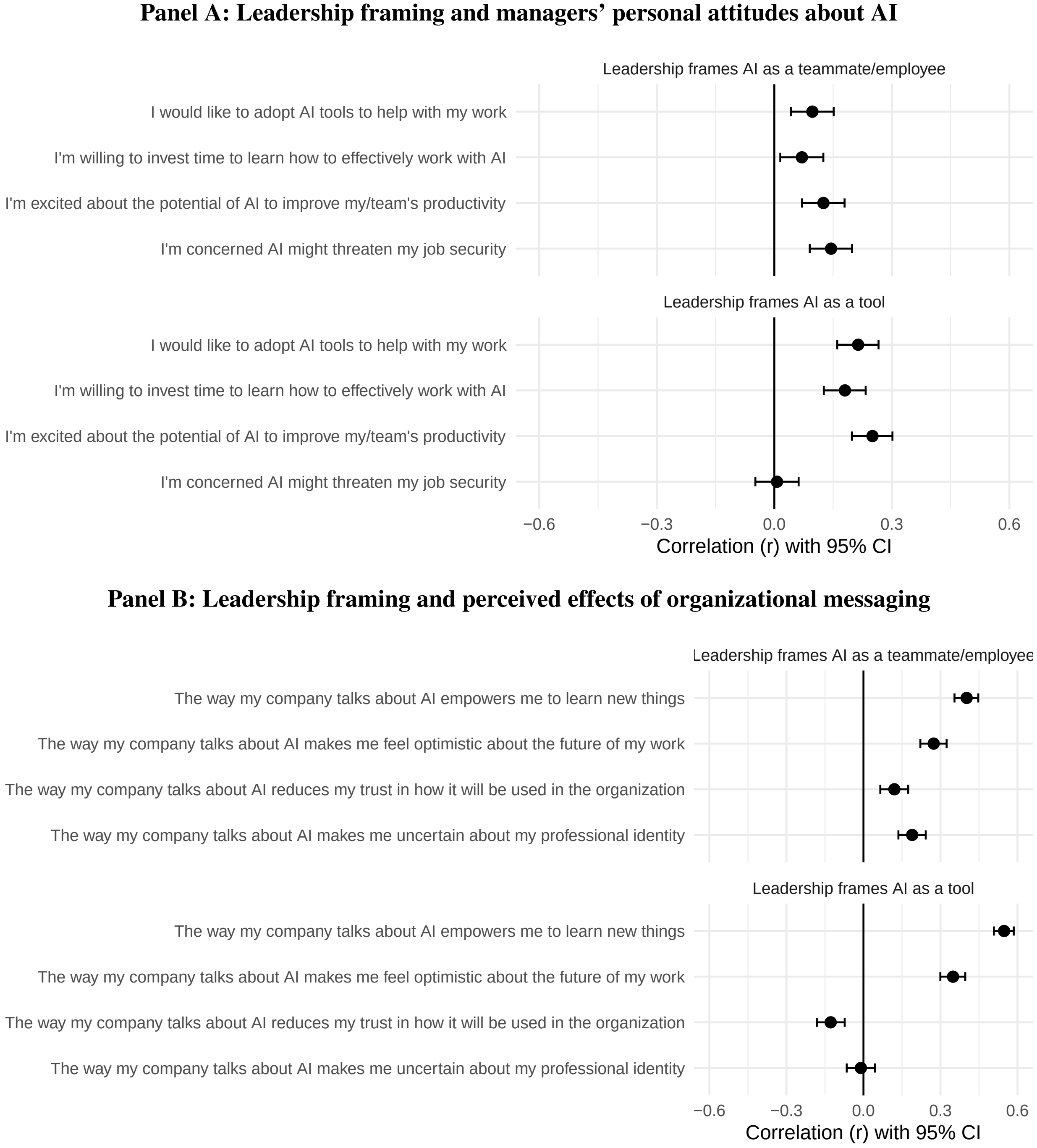
Table 2: Organizations With vs. Without AI Agents on the Org Chart

	No AI agent on org chart	Has AI agent on org chart
Panel A: Demographics		
HR	0.46	0.49
Finance	0.54	0.51
Company size: 0–100	0.25	0.15
Company size: 101–500	0.14	0.08
Company size: 501–5,000	0.32	0.30
Company size: 5,001–10,000	0.12	0.15
Company size: 10,000+	0.17	0.31
Fully remote	0.30	0.29
Hybrid	0.45	0.52
Fully in office	0.26	0.19
Manager	0.30	0.43
Director / VP	0.47	0.40
Executive	0.23	0.16
Industry: Technology	0.17	0.22
Industry: Healthcare	0.13	0.13
Industry: Financial services	0.16	0.19
Industry: Professional services	0.12	0.09
Industry: Manufacturing, construction	0.11	0.10
Industry: Other	0.31	0.26
Panel B: Organizations Positioning of AI		
Org frames AI as a tool	0.56	0.76
Org frames AI as a teammate	0.24	0.51
Org frames AI as a way to accelerate careers	0.27	0.51
Org has no clear AI stance	0.34	0.28
Org dissuades AI use	0.09	0.18
Direct manager encourages AI use	0.53	0.71
Panel C: Manager Positioning of AI		
Uses GenAI tools at least weekly	0.77	0.85
GenAI helps me do my job better	0.77	0.86

	No AI agent on org chart	Has AI agent on org chart
I review AI-generated content more thoroughly	0.59	0.66

Notes: This table compares organizations that do and do not place AI agents on their organizational charts. The survey question was "My company has AI agents listed on our org and/or work charts. (AI agents here means software or tools that take on tasks or roles — not people who work in AI-related roles.)" with the options of Yes/No/No answer. The first column provides the proportion of respondents who selected "No" who fall into each category. The second column is the proportion of respondents who selected "Yes."

Figure 4: Correlations between leadership framing of AI and managers' beliefs



Panel A: Leadership framing and managers' personal attitudes about AI

Panel B: Leadership framing and perceived effects of organizational messaging

Notes: All variables are in raw Likert responses, on a scale from 1 to 5. Panel A shows the correlations between leadership framing with managers' personal attitudes toward AI adoption and job security; Panel B shows the correlations between leadership framing with perceptions of how organizational AI messaging affects them. Points are pairwise correlations; bars are 95% confidence intervals.

Table 3: Summary Statistics across Treatment Arms

	Treatment arm				
	AI tool	Human employee	AI employee	N used	P value
Panel A: Balance Table for Registration Survey, $n = 1{,}261$					
N	420	419	422	1,261	
Female or Other (vs Male)	0.34	0.29	0.30	1,261	0.27
Age over 35	0.88	0.87	0.90	1,233	0.24
Graduate degree	0.61	0.61	0.64	1,253	0.57
US	0.87	0.83	0.85	1,261	0.18
Native English speaker	0.87	0.83	0.84	1,261	0.28
LLM use at least weekly	0.82	0.78	0.82	1,261	0.28
Remote or Hybrid	0.76	0.79	0.76	1,261	0.55
GenAI helps me do my job	0.80	0.82	0.82	1,230	0.79
Company size: 1,000+	0.56	0.56	0.52	1,254	0.40
Executive, Director, or VP	0.67	0.64	0.67	1,261	0.55
Stratifier: HR (vs Finance)	0.47	0.47	0.47	1,261	1.00
Stratifier: Frequent reviewer	0.71	0.71	0.71	1,261	0.99
Stratifier: LLM for work daily	0.40	0.41	0.41	1,261	0.99
Panel B: Participants that came back for the experiment, $n = 857$					
N	273	291	293	857	
Female or Other (vs Male)	0.33	0.30	0.29	857	0.59
Age over 35	0.87	0.85	0.89	837	0.44
Graduate degree	0.64	0.64	0.65	851	0.94
US	0.86	0.79	0.81	857	0.09
Native English speaker	0.84	0.80	0.79	857	0.34
LLM use at least weekly	0.84	0.78	0.82	857	0.15
Remote or Hybrid work arrangement	0.75	0.78	0.75	857	0.56
GenAI helps me do my job	0.81	0.81	0.82	843	0.89
Company size: Over 1,000 employees	0.57	0.56	0.48	851	0.07
Executive, Director, or VP	0.66	0.61	0.64	857	0.40
Stratifier: HR (vs Finance)	0.51	0.49	0.51	857	0.87
Stratifier: Frequent reviewer	0.70	0.71	0.73	857	0.66
Stratifier: Uses LLM for work daily	0.42	0.40	0.40	857	0.83
Panel C: Participants passing the second attention check, $n = 813$					
N	261	278	274	813	
Female or Other (vs Male)	0.33	0.30	0.30	813	0.60
Age over 35	0.87	0.85	0.88	793	0.49
Graduate degree	0.64	0.64	0.65	807	0.98
US	0.88	0.79	0.82	813	0.03
Native English speaker	0.85	0.81	0.81	813	0.39
LLM use at least weekly	0.84	0.77	0.82	813	0.08
Remote or Hybrid work arrangement	0.76	0.79	0.75	813	0.52
GenAI helps me do my job	0.80	0.80	0.83	800	0.71
Company size: Over 1,000 employees	0.56	0.56	0.47	807	0.06
Executive, Director, or VP	0.67	0.60	0.66	813	0.21
Stratifier: HR (vs Finance)	0.50	0.48	0.53	813	0.54
Stratifier: Frequent reviewer	0.70	0.72	0.75	813	0.47
Stratifier: Uses LLM for work daily	0.42	0.40	0.39	813	0.81

Table 4: Impact of Framing on Review Performance

	F1 Score	Precision	Recall	No submission
AI employee	0.015	−0.012	0.015	−0.007
	(0.020)	(0.021)	(0.022)	(0.029)
Human employee	0.000	−0.001	−0.002	0.016
	(0.018)	(0.020)	(0.020)	(0.028)
AI employee × AI on org chart	−0.084**	−0.065	−0.086**	−0.037
	(0.039)	(0.046)	(0.039)	(0.060)
AI on org chart	−0.025	−0.015	−0.030	0.057
	(0.022)	(0.025)	(0.023)	(0.036)
Num.Obs.	713	713	713	813
R2	0.144	0.092	0.134	0.058
Mean (omitted category)	0.490	0.707	0.423	0.095

Notes: This table reports treatment effects of the framing treatment on managers' performance outcomes, with those in the "AI Tool" condition as the omitted category, interacted with a binary indicator for if the manager reported having an AI agent on their company's org chart. All specifications include stratum fixed effects. All specifications include controls for the outcome at baseline. The sample includes all managers who passed an attention check. In Columns (1)-(3), the sample is narrowed to managers who submitted their reviewed documents. All standard errors are Huber–White robust.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Impact of Framing on Escalation Decisions

	(1)	(2)
AI employee	−0.042**	−0.031
	(0.021)	(0.047)
Human employee	−0.039**	−0.009
	(0.020)	(0.043)
AI employee × AI on org chart	0.057	0.251***
	(0.040)	(0.087)
AI on org chart	−0.017	0.016
	(0.025)	(0.052)
Num.Obs.	813	813
R2	0.018	0.023
Outcome	Escalate	Escalate
Costly escalation	No	Yes
Mean (omitted category)	0.975	0.448

Notes: This table reports treatment effects of the framing treatment on whether or not managers request an additional reviewer, with those in the "AI Tool" condition as the omitted category, interacted with a binary indicator for if the manager reported having an AI agent on their company's org chart. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. In Column (1), the outcome is 1 if the manager requested an additional reviewer. In Column (2) the additional review is costly for the managers. All standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

Table 6: Impact of Framing on Who is Accountable

	Me	My Team	AI System	Leadership
AI employee	0.523	−4.078**	1.701	0.793
	(2.880)	(1.986)	(1.738)	(0.998)
Human employee	−3.306	12.252***	−13.627***	2.315**
	(2.541)	(2.005)	(1.156)	(0.986)
AI employee × AI on org chart	−9.839*	6.835*	6.743*	−3.473
	(5.247)	(3.940)	(3.785)	(2.128)
AI on org chart	−4.941	−3.242	3.134**	3.988***
	(3.032)	(2.183)	(1.448)	(1.538)
Num.Obs.	813	813	813	813
R2	0.040	0.098	0.210	0.027
Mean (omitted category)	61.781	21.279	12.100	4.418

Notes: This table reports treatment effects of the framing treatment on managers' attribution of accountability across four possible actors: themselves ("Me"), their team, the AI system, and organizational leadership. The omitted category is the "AI Tool" condition. Treatment indicators for the employee framing (AI employee and Human employee) are interacted with a binary indicator for whether the manager reported having an AI agent on their company's organizational chart. Each column reports results from a separate OLS regression with the stated accountability outcome as the dependent variable. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

Table 7: Impact of Framing on Managers' Escalation Decision Alignment

	EDA (Obj)	EDA (Subj)
AI employee	-0.029	0.037
	(0.049)	(0.045)
Human employee	-0.016	0.050
	(0.046)	(0.040)
AI employee × AI on org chart	0.202**	-0.066
	(0.092)	(0.087)
AI on org chart	-0.076	-0.029
	(0.057)	(0.049)
Num.Obs.	713	813
R2	0.031	0.008
Mean (omitted category)	0.582	0.657

Notes: The table reports heterogeneous treatment effects on objective (EDA Obj) and subjective (EDA Subj) escalation decision alignment using OLS with stratum fixed effects, with the AI tool condition as the omitted category. Treatments are interacted with a binary indicator for whether AI agents are shown on the firm's organizational chart. The sample includes managers who passed an attention check. The mean of the omitted category is reported at the bottom of the table. All standard errors are Huber White robust.* p <0.1, ** p <0.05, *** p <0.01

Table 8: Impact of Framing on Delegation and Resource Allocation

	High Delegation	Invest in AI
AI employee	0.057	-0.010
	(0.045)	(0.033)
Human employee	0.250***	-0.004
	(0.046)	(0.033)
AI employee × AI on org chart	-0.006	0.011
	(0.101)	(0.073)
AI on org chart	0.032	-0.046
	(0.072)	(0.050)
Num.Obs.	807	768
R2	0.125	0.042
Mean (omitted category)	0.410	0.887

Notes: This table reports treatment effects of framing on managers' delegation and resource allocation decisions, with the AI tool condition as the omitted category. Treatments are interacted with a binary indicator for whether AI agents are shown on the firm's organizational chart. All regressions include stratum fixed effects. The sample includes managers who passed an attention check. The mean of the omitted category is reported at the bottom of the table. All standard errors are Huber White robust * p <0.1, ** p <0.05, *** p <0.01

Table 9: Impact of Framing on Source Attribution and Review Confidence

	Knew Author	Source attribution	Confidence
AI employee	0.111***	−0.228**	−0.122
	(0.036)	(0.101)	(0.102)
Human employee	−0.021	−0.232***	−0.204*
	(0.038)	(0.087)	(0.104)
AI employee × AI on org chart	0.086	0.192	0.011
	(0.056)	(0.196)	(0.217)
AI on org chart	0.010	−0.001	−0.084
	(0.045)	(0.100)	(0.157)
Num.Obs.	813	811	809
R2	0.036	0.016	0.069
Mean (AI tool)	0.776	3.642	3.463

Notes: This table reports treatment effects of framing on managers' beliefs about the documents and sign off confidence, with the AI tool condition as the reference category. "Knew Author" is a binary indicator for if the manager correctly reported who the author of the original documents was. "Source attribution" is a likert scale question on a scale of 1-5 asking the managers if knowing the source of the documents impacted how they reviewed them. "Confidence" is sign off confidence in the documents. Treatments are interacted with a binary indicator for whether AI agents are shown on the firm's organizational chart. All regressions include stratum fixed effects. The sample includes managers who passed an attention check. "Knew Author" is an indicator for whether the manager knew the content's author, while source attribution and confidence are measured on 1 to 5 scales. The mean of the AI tool condition is reported at the bottom of the table. All standard errors are Huber White robust.* p <0.1, ** p <0.05, *** p <0.01

Table 10: Impact of Framing on Binary Measures of Managers' AI Sentiments

	Excited	Want to Learn	Adoption	Job Insecurity
AI employee	−0.059*	−0.028	−0.035	−0.021
	(0.031)	(0.024)	(0.029)	(0.033)
Human employee	−0.033	−0.010	0.001	−0.037
	(0.026)	(0.021)	(0.025)	(0.031)
AI employee × AI on org chart	0.082	0.062	0.088	0.064
	(0.052)	(0.045)	(0.054)	(0.068)
AI on org chart	−0.008	−0.033	−0.033	−0.030
	(0.032)	(0.028)	(0.032)	(0.038)
Num.Obs.	813	813	813	813
R2	0.027	0.019	0.131	0.189
Mean (omitted category)	0.900	0.940	0.896	0.209

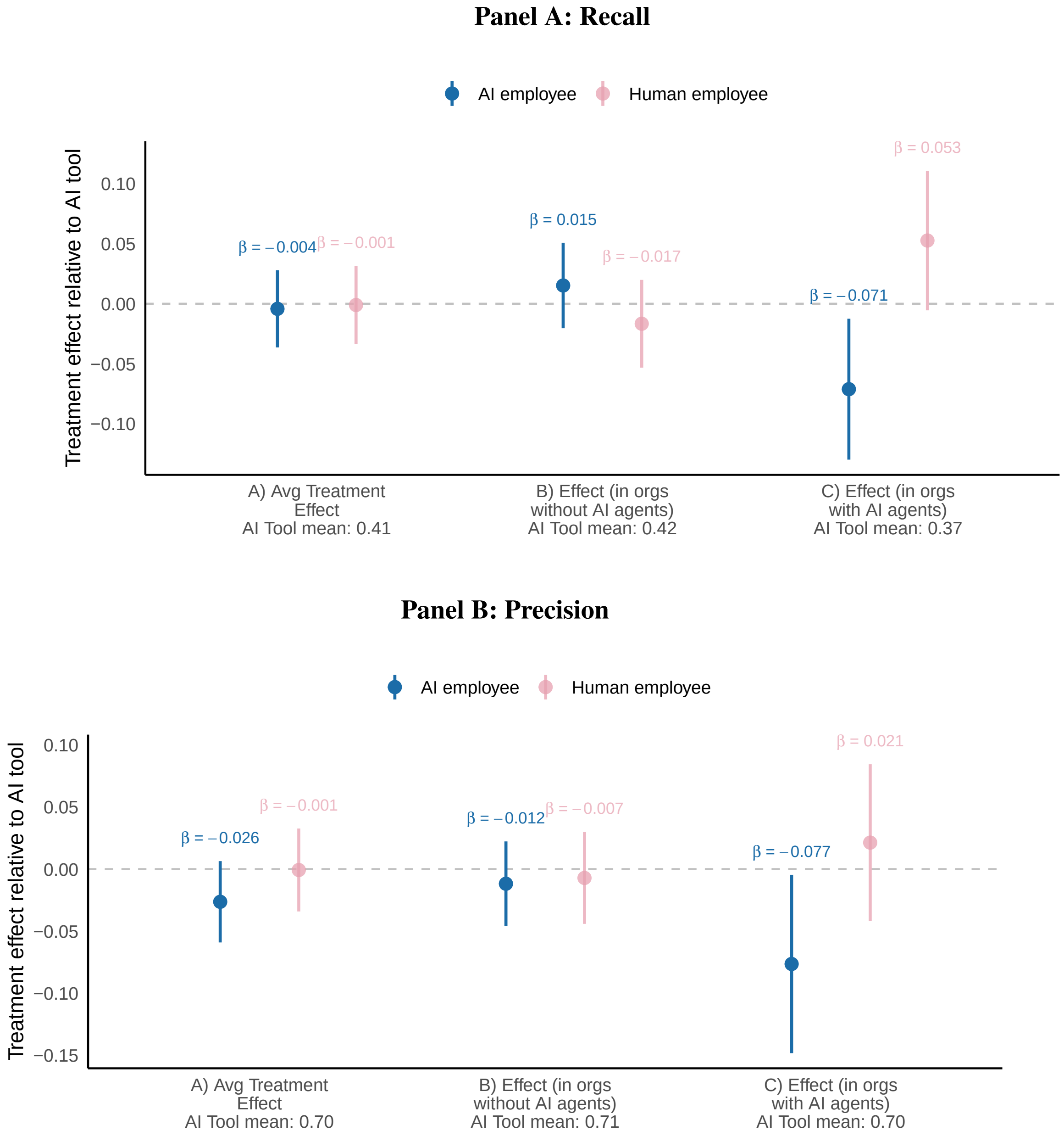
Notes: This table reports treatment effects of framing on binary measures of managers' AI related sentiments, with the AI tool condition as the omitted category. Treatments are interacted with a binary indicator for whether AI agents are shown on the firm's organizational chart. All regressions include stratum fixed effects. The sample includes managers who passed an attention check. Outcomes indicate whether the manager reports being excited about AI, wanting to learn more about AI, intending to adopt AI tools, or feeling job insecurity related to AI. The mean of the omitted category is reported at the bottom of the table. All standard errors are Huber White robust.* p <0.1, ** p <0.05, *** p <0.01

Table 11: Impact of Framing on Binary Measures of Comfort with AI Employees

	Comfort managing AI	Comfort with AI on org chart	Willingness to coach
AI employee	0.091**	0.047	0.019
	(0.044)	(0.045)	(0.033)
Human employee	0.053	0.068	0.084***
	(0.041)	(0.041)	(0.028)
AI employee × AI on org chart	0.098	0.118	0.036
	(0.079)	(0.087)	(0.062)
AI on org chart	0.007	0.086*	−0.047
	(0.049)	(0.050)	(0.035)
Num.Obs.	813	813	813
R2	0.070	0.048	0.028
Mean (omitted category)	0.577	0.328	0.846

*Notes:*This table reports treatment effects of framing on binary measures of managers' comfort with AI employees, with the AI tool condition as the omitted category. Treatments are interacted with a binary indicator for whether AI agents are shown on the firm's organizational chart. All regressions include stratum fixed effects. The sample includes managers who passed an attention check. Outcomes indicate whether managers report being comfortable managing an AI employee, comfortable with AI agents appearing on the organizational chart, and willing to coach or provide feedback to an AI employee. The mean of the omitted category is reported at the bottom of the table. All standard errors are Huber White robust.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 5: The Heterogeneous Impacts of AI Employee Framing on Precision and Recall



Panel A: Recall

Panel B: Precision

Notes: Points show estimated treatment effects relative to the AI-tool condition for the AI employee and human employee framings. Error bars show 90% confidence intervals with heteroskedasticity-robust standard errors. Estimates come from separate OLS regressions with stratum fixed effects, and each specification includes the other treatment arm as a main effect. Estimates come from OLS regressions with stratum fixed effects. "AI Tool mean" refers to the mean outcome in the AI-tool condition within each subgroup. For each outcome we control for the outcome at baseline to the regression. Regression output can be found in Table 4.

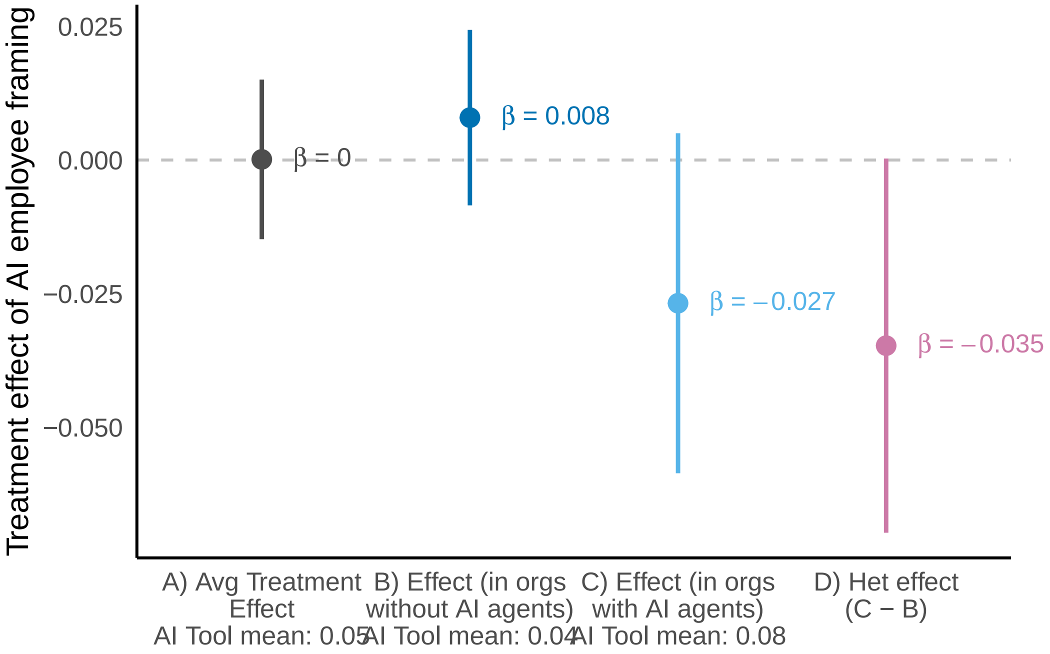


Figure 6: Accountability assigned to leadership

A.1 Average Treatment Effects

Table 12: Impact of Framing on Review Performance, Average Treatment Effects

	F1 Score	Precision	Recall	No submission
AI employee	-0.003	-0.026	-0.003	-0.016
	(0.019)	(0.020)	(0.020)	(0.027)
Human employee	-0.001	-0.002	-0.003	0.017
	(0.018)	(0.020)	(0.020)	(0.028)
Num.Obs.	713	713	713	813
R2	0.128	0.085	0.117	0.054
Mean (omitted category)	0.481	0.705	0.411	0.123

Notes: This table reports average treatment effects of the framing treatment on review performance outcomes. The dependent variables include standard classification performance metrics F1 score, precision, and recall, as well as an indicator for whether no review was submitted. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Impact of Framing on Escalation Decisions, Average Treatment Effects

	(1)	(2)
AI employee	−0.030	0.022
	(0.019)	(0.043)
Human employee	−0.039**	−0.009
	(0.020)	(0.043)
Num.Obs.	813	813
R2	0.016	0.007
Outcome	Escalate	Escalate
Costly escalation	No	Yes
Mean (omitted category)	0.966	0.46

Notes: This table reports average treatment effects of the framing treatment on managers' escalation decisions. The dependent variable is an indicator for whether the manager requested escalation. In Column (1), the outcome is 1 if the manager requested an additional reviewer. In Column (2) the additional review is costly for the managers. The omitted category is the "AI Tool" condition. Estimates are obtained from OLS regressions without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

Table 14: Impact of Framing on Delegation and Resource Allocation, Average Treatment Effects

	High Delegation	Invest in AI
AI employee	0.056	0.005
	(0.040)	(0.029)
Human employee	0.231^{***}	0.025
	(0.040)	(0.027)
Num.Obs.	807	768
R2	0.124	0.105
Mean (omitted category)	0.425	0.881

Notes: This table reports average treatment effects of the framing treatment on managers' delegation behavior and resource allocation decisions. The dependent variables include an indicator for high delegation and an indicator for willingness to invest in AI. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Impact of Framing on Managers' Escalation Decision Alignment, Average Treatment Effects

	EDA (Obj)	EDA (Subj)
AI employee	0.012	0.023
	(0.045)	(0.041)
Human employee	-0.017	0.050
	(0.046)	(0.040)
Num.Obs.	713	813
R2	0.024	0.005
Mean (omitted category)	0.576	0.655

Notes: This table reports average treatment effects of the framing treatment on managers' escalation decision alignment (EDA), measured using an objective benchmark (EDA (Obj)) and a subjective self-assessment measure (EDA (Subj)). The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check, with a smaller sample size for the objective EDA measure due to data availability. Standard errors are Huber–White robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Impact of Framing on Who is Accountable, Average Treatment Effects

	Me	My Team	AI System	Leadership
AI employee	−1.492	−2.587	3.086*	0.004
	(2.659)	(1.819)	(1.704)	(0.912)
Human employee	−3.362	12.228***	−13.591***	2.349**
	(2.546)	(2.007)	(1.159)	(0.998)
Num.Obs.	813	813	813	813
R2	0.023	0.094	0.189	0.013
Mean (omitted category)	59.238	21.241	13.724	5.291

Notes: This table reports average treatment effects of the framing treatment on managers' attribution of accountability across four possible actors: themselves ("Me"), their team, the AI system, and organizational leadership. The omitted category is the "AI Tool" condition. Estimates are obtained from OLS regressions without interaction terms, averaging treatment effects across managers regardless of whether they reported having an AI agent on their company's organizational chart. Each column reports results from a separate regression with the stated accountability outcome as the dependent variable. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

Table 17: Impact of Framing on Binary Measures of Comfort with AI Employees, Average Treatment Effects

	Comfort managing AI	Comfort with AI on org chart	Willingness to coach
AI employee	0.111***	0.070*	0.027
	(0.040)	(0.042)	(0.030)
Human employee	0.053	0.069*	0.084***
	(0.041)	(0.041)	(0.028)
Num.Obs.	813	813	813
R2	0.067	0.035	0.026
Mean (omitted category)	0.582	0.352	0.839

Notes: This table reports average treatment effects of the framing treatment on three binary measures of managers' comfort and willingness to engage with AI employees: comfort managing an AI employee, comfort with having an AI agent on the organizational chart, and willingness to coach an AI employee. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Impact of Framing on Binary Measures of Managers' AI Sentiments, Average Treatment Effects

	Excited	Want to Learn	Adoption	Job Insecurity
AI employee	-0.042	-0.014	-0.016	-0.007
	(0.027)	(0.021)	(0.026)	(0.031)
Human employee	-0.033	-0.010	0.001	-0.038
	(0.026)	(0.021)	(0.025)	(0.031)
Num.Obs.	813	813	813	813
R2	0.024	0.017	0.129	0.188
Mean (omitted category)	0.912	0.939	0.893	0.211

Notes: This table reports average treatment effects of the framing treatment on four binary measures of managers' AI-related sentiments: excitement about AI, interest in learning more about AI, willingness to adopt AI tools, and perceived job insecurity due to AI. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

Table 19: Impact of Framing on Source Attribution and Review Confidence, Average Treatment Effects

	Knew Author	Source attribution	Confidence
AI employee	0.129***	−0.187**	−0.020
	(0.032)	(0.092)	(0.043)
Human employee	−0.020	−0.232***	−0.027
	(0.038)	(0.087)	(0.042)
Num.Obs.	813	811	813
R2	0.032	0.014	0.053
Mean (AI tool)	0.762	3.667	3.460

Notes: This table reports average treatment effects of the framing treatment on managers' perceptions of review sources and confidence in their evaluations. The dependent variables include an indicator for whether the manager reported knowing the author of the content, a measure of source attribution, and a self-reported confidence measure. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Impact of Framing on Delegation and Resource Allocation, Average Treatment Effects

	High Delegation	Invest in AI
AI employee	0.056	0.005
	(0.040)	(0.029)
Human employee	0.231***	0.025
	(0.040)	(0.027)
Num.Obs.	807	768
R2	0.124	0.105
Mean (omitted category)	0.425	0.881

Notes: This table reports average treatment effects of the framing treatment on managers' delegation behavior and resource allocation decisions. The dependent variables include an indicator for high delegation and an indicator for willingness to invest in AI. The omitted category is the "AI Tool" condition. Estimates are obtained from separate OLS regressions for each outcome without interaction terms. All regressions include stratum fixed effects. The sample includes all managers who passed an attention check. Standard errors are Huber–White robust. * p <0.1, ** p <0.05, *** p <0.01

A.2 Do the AI Employee effects differ from the Human employee effects?

In order to assess whether the results reflect generic employee framing rather than the specific AI employee framing, we re-estimate the heterogeneity specification using the Human Employee condition as the omitted category.

$$\begin{aligned} y_i = \beta_0 + \beta_1 \mathbb{1}(\text{AI Emp}_i) + \beta_2 \mathbb{1}(\text{Human Emp}_i) + \beta_3 \text{OrgAgents}_i \\ + \beta_4 \left(\mathbb{1}(\text{AI Emp}_i) \times \text{OrgAgents}_i\right) + \gamma \tilde{y}_i^{pre} + \delta M_i^{miss} + \mathbf{X}_i \Pi + \varepsilon_i. \end{aligned} \tag{3}$$

Here, OrgAgents$_i$ is an indicator equal to 1 if the participant reported in the registration survey that their organization includes AI agents on organizational or work charts. In this specification, β_4 captures whether the effect of framing the upstream producer as an AI employee rather than a human employee differs for managers in organizations that already formalize AI agents. Table 21 reports the results.

Table 21: Impact of Framing on Main Outcomes, Human Employee Active-Control

	F1 Score	Escalate	Accountability (me)	Accountability (AI)	Accountability (team)
AI employee	0.015	−0.021	3.829	15.328***	−16.329***
	(0.020)	(0.047)	(2.716)	(1.341)	(2.053)
AI tool	0.000	0.009	3.306	13.627***	−12.252***
	(0.018)	(0.043)	(2.541)	(1.156)	(2.005)
AI employee × AI on org chart	−0.084**	0.251***	−9.839*	6.743*	6.835*
	(0.039)	(0.087)	(5.247)	(3.785)	(3.940)
AI on org chart	−0.025	0.016	−4.941	3.134**	−3.242
	(0.022)	(0.052)	(3.032)	(1.448)	(2.183)
Num.Obs.	713	813	813	813	813
R2	0.144	0.023	0.040	0.210	0.098
Mean (omitted category)	0.481	0.451	56.108	0.000	34.826

Notes: This table reports treatment effects of the framing treatment on managers' performance outcomes, with those in the "Human Employee" condition as the omitted category, interacted with a binary indicator for if the manager reported having an AI agent on their company's org chart. All specifications include stratum fixed effects. All specifications include controls for the outcome at baseline. The sample includes all managers who passed an attention check. In Columns (1) the sample is narrowed to managers who submitted their reviewed documents. All standard errors are Huber–White robust.* p <0.1, ** p <0.05, *** p <0.01

A.3 Robustness Checks

Table 22: Impact of Framing on Review Performance, by Manager AI Use

	F1 Score	Additional Review	Accountability	Accountability
			Self	AI System
AI employee	−0.012	0.063	−2.717	3.067
	(0.024)	(0.054)	(3.174)	(2.057)
Human employee	0.000	−0.009	−3.108	−13.679***
	(0.019)	(0.043)	(2.568)	(1.169)
AI employee × Daily AI Use	0.034	−0.096	3.633	0.090
	(0.033)	(0.076)	(4.576)	(2.825)
DailyAI Use	0.015	0.053	2.330	−1.333
	(0.019)	(0.044)	(2.621)	(1.129)
Num.Obs.	709	808	808	808
R2	0.111	0.006	0.018	0.187
Mean (omitted category)	0.478	0.453	58.642	14.608

Notes: This table reports treatment effects of the framing treatment on managers' performance outcomes, with those in the "AI Tool" condition as the omitted category, interacted with a binary indicator for if the manager reported using AI on a daily basis. Because "Daily AI use" was one of the variables in the stratified randomization, we cannot include stratum fixed effects, so instead control for the other variables stratified on. F1 specification includes controls for the outcome at baseline. The sample includes all managers who passed an attention check. In Columns (1), the sample is narrowed to managers who submitted their reviewed documents. All standard errors are Huber–White robust.* p <0.1, ** p <0.05, *** p <0.01

Table 23: Impact of Framing on Review Performance, by Tech Industry

	F1 Score	Additional Review	Accountability	Accountability
			Self	AI System
AI employee	-0.002	0.030	-0.767	2.568
	(0.020)	(0.046)	(2.851)	(1.834)
Human employee	-0.003	-0.008	-3.359	-13.597***
	(0.018)	(0.043)	(2.556)	(1.150)
AI employee × Tech Industry	0.010	-0.055	-4.004	3.225
	(0.044)	(0.099)	(5.567)	(3.397)
Tech Industry	0.033	-0.059	0.650	-1.144
	(0.024)	(0.055)	(3.213)	(1.183)
Num.Obs.	713	813	813	813
R2	0.109	0.007	0.015	0.185
Mean (omitted category)	0.479	0.475	58.700	14.175

Notes: This table reports treatment effects of the framing treatment on managers' performance outcomes, with those in the "AI Tool" condition as the omitted category, interacted with a binary indicator for if the manager reported being employed by a firm in the technology industry (17% of managers.) All specifications include stratum FE and the F1 specification includes controls for the outcome at baseline. The sample includes all managers who passed an attention check. In Columns (1), the sample is narrowed to managers who submitted their reviewed documents. All standard errors are Huber–White robust.* p <0.1, ** p <0.05, *** p <0.01

Table 24: Impact of Framing on Review Performance, by Firm Size

	F1 Score	Additional Review	Accountability	Accountability
			Self	AI System
AI employee	0.005	0.001	−5.785*	5.544***
	(0.024)	(0.057)	(3.323)	(2.001)
Human employee	−0.002	−0.011	−3.364	−13.622***
	(0.019)	(0.043)	(2.545)	(1.150)
AI employee × Large Firm	−0.010	0.032	8.380*	−4.787*
	(0.033)	(0.075)	(4.435)	(2.731)
Large Firm (1,000+ employees)	0.005	−0.051	−4.866*	2.255**
	(0.019)	(0.043)	(2.542)	(1.087)
Num.Obs.	713	813	813	813
R2	0.105	0.005	0.019	0.188
Mean (omitted category)	0.489	0.483	63.198	10.905

Notes: This table reports treatment effects of the framing treatment on managers' performance outcomes, with those in the "AI Tool" condition as the omitted category, interacted with a binary indicator for if the manager reported being employed by a firm with more than 1,000 employees (52% of managers.) All specifications include stratum FE and the F1 specification includes controls for the outcome at baseline. The sample includes all managers who passed an attention check. In Columns (1), the sample is narrowed to managers who submitted their reviewed documents. All standard errors are Huber–White robust.* p <0.1, ** p <0.05, *** p <0.01

B Methods: Details

B.1 Outcomes

B.1.1 Error detection performance

Our primary performance outcome is micro-averaged $F1$ for error detection aggregated across all documents reviewed by participant i. Let:

- TP_i: number of errors correctly flagged,
- FP_i: number of non-errors flagged,
- FN_i: number of errors not flagged.

Precision and recall are:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}, \qquad \text{Recall}_i = \frac{TP_i}{TP_i + FN_i}.$$

The $F1$ score is:

$$F1_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}.$$

If a participant made no flags, we set F1 to missing.

B.1.2 Escalation

Request for additional review. We measure whether participants request another reviewer before finalization two ways. First, we ask them if they would like the documents to be additionally reviewed “Would you like to ask someone to conduct additional review of the documents you submitted?” (Yes/No/No answer). Because we believed most would say yes, we ask a second version with stakes, requesting review was rewarded when performance was low and penalized when performance was high, reducing purely expressive escalation. In the second question we added the following note “If you select ’Yes’ and caught fewer than 50% of errors, you’ll gain 3 tickets (good job recognizing uncertainty!). If you caught more than 50%, you’ll lose 3 tickets (unnecessary review). If you select ’No,’ you keep current tickets.”

B.1.3 Sign-off confidence.

Participants reported confidence in signing off on the reviewed materials on a 1–5 Likert scale ($\textit{confidence}_i$). We additionally define:

$$\textit{HighConf}_i = \mathbb{1}\{\textit{confidence}_i \in \{4, 5\}\}.$$

B.1.4 Governance and resource allocation preferences

Delegation of decision rights. Participants recommended a governance structure for deploying an assistant like the one they worked with, ranging from autonomy to “assistance only” to “do not use.” We define:

$$\textit{HighDelegation}_i = 1$$

if the participant recommends full or partial decision authority for the assistant, and 0 otherwise.

Resource allocation preference. Participants chose between hiring additional employee(s) versus investing the same amount in improving/integrating the AI system. We define:

$$\textit{InvestInAI}_i = 1$$

if the participant recommends investing in the AI system, and 0 if they recommend hiring.

B.1.5 Post-task attitudes

Key attitude outcomes include adoption intent, excitement about AI, willingness to invest time to learn AI skills, and job insecurity, each measured on 1–5 Likert scales and also coded as high-endorsement indicators:

$$\textit{HighAdoption}_i = \mathbb{1}\{\textit{adopt}_i \in \{4,5\}\}, \qquad \textit{HighInsecurity}_i = \mathbb{1}\{\textit{insecurity}_i \in \{4,5\}\}$$

$$\textit{HighExcited}_i = \mathbb{1}\{\textit{excited\_prod}_i \in \{4,5\}\}, \qquad \textit{HighInvestTime}_i = \mathbb{1}\{\textit{invest\_time}_i \in \{4,5\}\}$$

B.1.6 Attention checks

Participants reported who they believed initially drafted the documents (AI tool vs. human employee vs. AI employee), used to assess whether framing was received as intended. We do not filter the sample on this, instead we use it as an outcome.

A second attention check asked the participants ”Please answer number 3” during the block of likert scale questions where the possible answers were between 1 and 5.