How Do Organizations Learn? The Diffusion of Scientific Evidence on Generative AI*

Mahvish Shaukat† *World Bank*

Andreas Stegmann‡ *University of Warwick*

Mattie Toma§ *University of Warwick*

June 2025

Abstract

We study the dissemination of scientific evidence in a large organization through a field experiment at the World Bank. We experimentally vary whether research findings on the impacts of generative AI are shared with senior or junior staff, and vary beliefs about peer adoption and evidence credibility. Providing evidence to senior staff significantly increases transmission and diffusion as measured by self-reports, engagement with study materials, and colleagues' recall of study details. In contrast, changing beliefs about peer adoption or credibility has no detectable effects. The results highlight the importance of organizational hierarchy in shaping informal information flows.

JEL Classifications: D83, D73, M50, C93

Keywords: Information diffusion, Organizational learning, Evidence use, Generative AI, Field experiments

*We are grateful to Shruti Agarwal, Abdullah Mahmood, and Matthew Tan for excellent research assistance. We thank Anil Doshi, John Firth, Xavier Gine, Ryan Hahn, Oliver Hauser, Erin Kelley, Adnan Q. Khan, David McKenzie, Shakked Noy, Gautam Rao, Daniel Rogger, Chris Roth, Sergio Schmukler, Ashish Shenoy, Niharika Singh, Adam Szeidl, Michael Thaler, Gabriel Tourek, as well as numerous conference and seminar participants for helpful feedback. We thank University of Warwick's Experimental and Behavioral Economics Research fund, the Warwick Business School Primary Data Support Fund, and DECRG Research Support for their generous funding. This study is pre-registered on the AEA Social Science Registry (AEARCTR-0012551), and received data collection approval from the World Bank Data Privacy Office as well as ethics approval through the University of Warwick (Round 1: HSSREC 105.22-23; Round 2: HSSREC 234/24-25; Prediction study: HSSREC 47.24-25). The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

†World Bank Development Research Group. Email: mshaukat@worldbank.org.

‡Department of Economics. Email: andreas.stegmann@warwick.ac.uk.

§Warwick Business School. Email: mattie.toma@warwick.ac.uk.

1 Introduction

The success of a new idea or technology in an organization often depends on how effectively information flows to the people who need it. In many organizations, decisions are made through a hierarchy, and communication must move up and down across different levels. A frontline worker may never hear about the new idea, or a decision-maker may remain unaware of its potential. Each layer can slow down, filter, or distort the message. Additional information frictions may arise from individual-level factors, such as employees' beliefs about the relevance or credibility of the information. Together, these factors shape organizational learning, that is whether the idea is transmitted and taken up (Atkin et al., 2017; Bloom et al., 2013; Buera and Lucas, 2018). This is particularly consequential when considering whether new scientific evidence is shared in domains like health, education, technology, or environmental policy, where whether decision-makers learn about and act on the information can affect the well-being of large populations (DellaVigna et al., 2024; Ruggeri, 2024; Slavin, 2002).

In this paper, we leverage a large-scale field experiment to study what determines how scientific evidence is transmitted within a major organization in economic development, the World Bank. We focus on the dissemination of evidence on the impacts of generative AI (genAI) during the early stages of adoption of the new technology in workplaces across the world. We assigned all 511 divisions within the World Bank's headquarters to treatment and control conditions. In treated divisions, we shared scientific evidence about the impacts of genAI in the workplace—as estimated in two recent academic studies—with one randomly selected employee during an online survey. To understand the role of organizational rank in shaping information diffusion, we randomly varied whether we shared evidence with a senior or junior employee in each division. We also included two cross-randomized treatments to explore the role of beliefs in motivating evidence sharing: We randomly provided additional information on *i)* colleagues' adoption of and attitudes towards genAI, and *ii)* the credibility of the study findings. In control divisions, we surveyed one randomly selected employee but did not provide any evidence. We refer to this initial group of employees in treated and control divisions as *seeds*. After treated seeds completed the online survey, we shared links to the study and an infographic summarizing the findings.

We measured the transmission of evidence in three ways: First, we elicited seeds' self-reported intentions to share the evidence. Second, we recorded the total number of clicks on the links shared with seeds to measure actual engagement with the evidence among seeds and their colleagues. Finally, in a separate survey fielded 2 to 3 weeks after the initial dissemination of the evidence, we assess seeds' colleagues' recall of details about the studies we had shared with seeds.

We implemented two rounds of the experiment between December 2023 and April 2024. In the first round, we shared evidence from Noy and Zhang (2023) on the impacts of ChatGPT on *productivity* at the workplace, while in the second round, we shared evidence from Hauser and Doshi (2024) on the impacts of ChatGPT on *creativity*.[1] The two rounds of our experiment provide insights into how evidence on different aspects of the new technology influences transmission.

[1]Both studies are online experiments that randomly vary whether skilled professionals have access to ChatGPT when completing writing tasks. The authors document substantial improvements in the quality of the output across several dimensions of productivity and creativity. Further details are provided in Section 2.3.

Our first set of results indicates that sharing evidence with a senior seed increases information diffusion. Senior seeds are 14.2 percentage points more likely to report that they are "likely" or "very likely" to share the evidence with colleagues ($p = 0.01$) compared to junior seeds. In divisions where the evidence was shared with a senior seed, the likelihood of multiple clicks per infographic link increased by 15.8 percentage points ($p < 0.01$). We also find that senior seeds' colleagues are 5.0 percentage points more likely to recall at least one notable study feature ($p = 0.03$) relative to junior seeds' colleagues.

Our second set of results indicate that seeds' beliefs about their colleagues' adoption of and attitudes toward genAI do not play a major role in the diffusion of scientific evidence on the workplace impacts of ChatGPT. To change seeds' beliefs, we implemented a baseline survey of colleagues' behavior and beliefs about ChatGPT. Then, we randomly assigned seeds to see different subsets of the data: half of the seeds saw more positive information about their colleagues' adoption and views of genAI, while another half were exposed to more negative information.[2] The information treatment is motivated by a conformity-based framework (Bursztyn et al., 2020; Bursztyn and Yang, 2022): individuals may use peers' behavior as a cue for social acceptance, possibly shaping incentives to communicate. Seeds informed that ChatGPT adoption among colleagues is low may perceive evidence on the technology's benefits as less acceptable to share. In contrast, those told that adoption is high may view the same evidence as more valuable to share, anticipating that it aligns with prevailing norms and will be more positively received by colleagues. While our information treatments had a substantial impact on seeds' second-order beliefs about their colleagues' ChatGPT adoption and perceptions of its impacts ($p < 0.01$), we find no evidence supporting this hypothesis.

Our final set of results suggests sharing information on the evidence's credibility does not substantially affect seeds' decisions to share the evidence with colleagues. Seeds in the first round of the experiment who were informed that the study was published in *Science*, conducted by MIT-affiliated researchers, and independently replicated rated the evidence as suggestively more credible ($p = 0.11$).[3] However, we find no significant differences in sharing intentions or colleague recall, and a directionally negative effect on actual engagement with the materials provided to seeds.

We survey a sample of experts, including researchers at the World Bank, and find they mispredict the relative effects of our treatments. Experts expect our treatment altering second-order beliefs about colleagues' adoption and attitudes toward ChatGPT to be most impactful, while they expect hierarchy will be the least relevant predictor of evidence sharing among our treatments.

Across our treated divisions, we document overall information diffusion is limited, suggesting frictions in the diffusion of evidence within the organization. In our contact survey, only 3% of seeds' colleagues recall any relevant aspect of the evidence shared with seeds approximately 2 to 3 weeks after we initially shared the evidence in round 1. We measure no recall in round 2. This is despite a majority

[2]This approach of exogenously altering beliefs has been used in both lab and field settings: Eil and Rao (2011) assign people to "good" or "bad" news based on their rank in comparison to randomly-selected peers. Bottan and Perez-Truglia (2025) randomly assign the source of information on the evolution of home prices.

[3]Our credibility treatment is motivated by models such as Gentzkow and Shapiro (2006), which formalize how information recipients use signals of source quality to update their beliefs. Fong (2020) show policymakers rely on trusted sources when information is incomplete. We test whether emphasizing the scientific rigor and institutional credibility of evidence increases its diffusion.

of seeds expressing an intention to share and substantial engagement with the materials shared. The limited diffusion is striking—particularly given the salience of genAI at the time of our experiment.

The finding that seniority is predictive of sharing intentions, actual engagement with evidence, and colleagues' recall may reflect both selection into senior ranks and the causal effects of holding a senior position. Survey data we collect suggests that senior seeds place more weight on the value of the evidence to their colleagues and less on potential interpersonal frictions when deciding whether to share evidence. Senior seeds also differ systematically from junior seeds, being, on average, more educated and more likely to be male, pointing to a role for selection. Notably, senior seeds are also more likely to have managerial responsibilities and may view evidence dissemination as part of their professional mandate. Consistent with this, senior seeds with managerial responsibilities are indeed more likely to state high sharing intentions. However, the effects of seniority persist even when excluding seeds with managerial responsibilities from our analysis, suggesting senior seeds share evidence for reasons beyond their formal managerial duties.

A key feature of our experimental design is the full cross-randomization of the information treatments and the seniority treatment. This allows us to test whether junior seeds' sharing decisions are influenced by pessimistic beliefs about the technology's acceptability or concerns about the credibility of the information. While the estimates are imprecise, we do not find statistically significant effects of either factor on sharing behavior.

Our work contributes to a literature in organizational economics that examines barriers to communication and their wide-ranging implications for firms, including effects on technology adoption, coordination, innovation, and productivity (Atkin et al., 2017; Garicano and Rayo, 2016; Levitt and Neckermann, 2015; Sandvik et al., 2020; Menzel, 2021; Cai and Wang, 2022; Guillouet et al., 2024; Macchiavello and Morjaria, 2015; Juhasz et al., 2024). Besley et al. (2022) highlight the limited evidence on how organizations, including bureaucracies, acquire and transmit knowledge. Our study is one of a few to provide causal evidence from a field experiment inside a large bureaucratic organization.

We address one specific and practical question: how should initial "seeds" be selected to maximize information diffusion? In particular, we examine the hierarchical rank of seeded individuals, leveraging data on the World Bank's organizational structure to compare diffusion patterns for senior versus junior staff.[4] We further contribute to this literature by studying the diffusion of scientific evidence and the role of beliefs — both about colleagues' receptivity and about the credibility of the information — as a potential mechanism shaping communication dynamics within organizations.

Our study also contributes to the rapidly growing literature on the use of evidence in policy-making (Hjort et al., 2021; DellaVigna et al., 2024; Rogger and Somani, 2023; Rey Beil et al., 2025; Rao, 2024; Crowley et al., 2021). We add to this work by experimentally examining the conditions under which research *disseminates* within an organization that consults for, and routinely provides technical assistance

[4]The question of whom to seed with information to maximize diffusion has been studied in informal settings. In development economics, scholars have examined social learning and the diffusion of agricultural technologies (Beaman et al., 2021; BenYishay, 2019; Conley and Udry, 2010; Foster and Rosenzweig, 1995; Hardy and McCasland, 2021; Magruder, 2018; Duflo et al., 2023). Related studies analyze information diffusion through social networks in village contexts, with applications to microfinance (Banerjee et al., 2013, 2019; Breza and Chandrasekhar, 2019; Breza, 2016), and online, particularly in the context of health behaviors (Alatas et al., 2024). In contrast, relatively little is known about how diffusion operates within formal organizations.

to, policymakers around the world. By focusing on dissemination within a key intermediary institution — an essential link in the chain that delivers research to policymakers — we also speak to a related literature on the communication of scientific evidence and the formation of beliefs by stakeholders along the evidence-to-policy pipeline (Thaler et al., 2025; Vivalt and Coville, 2023).

We study these questions in the context of genAI adoption. Existing work in economics has mostly focused on the consequences of AI for labor markets and productivity (Acemoglu, 2025; Acemoglu et al., 2022; Crafts, 2021; Frank et al., 2019; Beraja et al., 2023b,a; Kreitmeir and Raschky, 2023; Humlum and Vestergaard, 2025). We contribute to a nascent literature on the dynamics that shape how individuals within an organization change their beliefs about or adopt AI (Goldfarb et al., 2020; Mehmood et al., 2024). We also present novel descriptive evidence on prior beliefs about, and workplace adoption of, genAI (Bick et al., 2024).

2 Setting and Design

2.1 Setting

Our pre-registered field experiment was conducted at the World Bank's headquarters in Washington, D.C. The study population encompassed 6,254 employees working in 511 divisions across the organization. Evidence use at the World Bank is widespread: 45% of World Bank employees report seeking out research evidence at least weekly, while only 17% say they never do.

2.2 Sampling and Recruitment Procedures

We implemented the experiment twice between December 2023 and April 2024. In each round, we randomly selected one employee per division and invited them to complete our survey, introduced as a brief study on AI use at the World Bank.[5] Each selected employee was notified that our aim was to survey one individual per division and had three days to respond. Non-respondents were replaced by another randomly selected employee (provided there were additional eligible employees within the division), up to three times per division. We refer to the selected employees as *seeds*. In total, 285 seeds in round 1 and 176 seeds in round 2 completed the seed survey, for a division-level response rate of 56% in round 1 and 38% in round 2.

We then contacted seeds' colleagues 2 to 3 weeks after the final round of seed surveys. We refer to these employees as *contacts*. We recruited 609 contacts in round 1 (397 of these contacts worked in divisions in which a seed completed a survey in round 1) and 461 contacts in round 2, corresponding to a response rate of 12% in round 1 and 17% in round 2.

2.3 Evidence on the Impacts of Generative AI

In each round, we shared evidence on impacts of ChatGPT in the workplace with seeds.

[5] Appendix Figure C.1 shows example recruitment emails.

In round 1 of the experiment, we shared evidence from Noy and Zhang (2023), one of the first experimental studies to examine the productivity effects of ChatGPT on skilled workers. The study, published in *Science* four months prior to the launch of our experiment, found that access to ChatGPT significantly improved task completion time and quality of typical writing tasks within a sample of skilled professionals. Treated seeds are informed about the average treatment effects for these two outcomes.

In round 2, we shared evidence from Hauser and Doshi (2024), which examines ChatGPT's impact on creativity using a similar experimental design. Participants were tasked with writing a short story, simulating an open creativity assignment. The study found that ChatGPT access significantly increased third-party assessments of story novelty and usefulness. Analogous to round 1, treated seeds are informed about the average treatment effects on these two outcomes. While unpublished at the time of our experiment, the article has since been published in *Science Advances*.

The articles shared in each round have several similarities: Both are online experiments with comparable samples of professionals, evaluating the effects of ChatGPT access on writing tasks. The studies' main outcomes are based on output produced by study participants, though each study has a different focus as outlined above.

Notably, we measure different priors about the evidence across rounds: respondents expect ChatGPT to improve workplace productivity more than creativity ($p < 0.01$). This allows us to study how prior beliefs on different aspects of the technology may shape information diffusion.

2.4 Experimental Design

Our randomization is implemented at the division level. Appendix Figure A.1 shows the four levels of randomization in round 1. Treatment assignments were stratified by division size. All four levels of randomization are described below. In round 2, we maintained the overall treatment status of divisions from round 1 and re-randomized the seniority and information conditions while excluding the credibility condition.[6]

We randomly assigned 75 percent of divisions to treatment, where a randomly selected seed received the empirical evidence on the impact of ChatGPT in the workplace described in Section 2.3. The remaining 25 percent formed the control group, with seeds in these divisions receiving no evidence on ChatGPT's effects.

To measure spillover effects, we used a saturation design. The World Bank's 42 Program Management Units (PMUs), each comprising multiple divisions, were randomly assigned to one of three saturation levels: 50%, 75%, or 100% of divisions assigned to treatment.

We cross-randomized seeds to the following sub-treatments:

Organizational rank We randomly varied whether we invited a senior (World Bank grades GG or above, i.e. positions such as Senior Specialist, Senior Economist, or Director) or junior (World Bank grades GF or below, i.e. positions such as Economist, Analyst, or Assistant) seed to complete the survey.

[6]Other differences between rounds were minimal; the procedures for recruiting survey participants remained broadly consistent in both stages of the experiment. Further details are in Online Appendix C.

Information about ChatGPT adoption We experimentally altered seeds' beliefs about *(i)* colleagues' adoption of generative AI and *(ii)* their views on whether World Bank employees should adopt ChatGPT. Using an active control group design, we randomly assigned each seed to receive either positive or negative information about their colleagues' views.[7]

To generate this information, we conducted a pre-survey among 72 World Bank employees not eligible for the main experiment,[8] asking about their ChatGPT use and the degree to which they agree that World Bank employees should frequently use ChatGPT. We computed summary statistics for two subsets of 40 respondents and shared them with the respective treatment groups. In the *positive* information condition, seeds learned that 75% of colleagues in the subset use ChatGPT, and 80% agree or strongly agree that it should be used frequently. In the *negative* information condition, these figures were 20% and 18%, respectively.[9] Seeds knew they were learning only about a subset of pre-survey results. After both rounds of the experiment, we emailed all seeds with the full sample statistics to correct beliefs.

Credibility We explore whether enhancing a study's perceived credibility influences information diffusion. In round 1 of the experiment, seeds in the credibility treatment receive three additional details: the paper was published in *Science*, its authors are affiliated with MIT, and its findings were independently replicated by researchers from Harvard Business School and Wharton. Seeds not assigned to the credibility treatment receive no additional information within the online survey.[10] We do not vary the credibility of the study in round 2 as Hauser and Doshi (2024) was not published at the time.

2.5 Data Collection

Seed Survey The seed survey began with questions on general attitudes towards ChatGPT. Treated seeds were then provided the *positive* or *negative* information from our pre-survey about their colleagues' adoption and perception towards ChatGPT, and all seeds were asked to predict the share of colleagues using ChatGPT and those who somewhat or strongly agree it should be used often. 10% of respondents were randomly selected to receive a bonus payment if one of their guesses was within 5 percentage points of the true share.

Next, treated seeds received a brief description of the study design in Noy and Zhang (2023) or Hauser and Doshi (2024) (without revealing results) and were asked to predict two key findings.[11] Seeds also guessed their colleagues' "typical prediction" for both questions, providing us with their beliefs about colleagues' priors about the evidence. Seeds were again paid bonuses based on the accuracy of their predictions.

[7] Active control group designs help ensure treatment effects reflect genuine belief updating rather than experimenter priming (Haaland et al. (2023)).

[8] Employees outside headquarters and those in Human Resources.

[9] Appendix B presents descriptive statistics on ChatGPT use at the World Bank.

[10] However, they may have learned about the information post-survey, potentially attenuating treatment effects for outcomes measured after the survey experiment.

[11] In round 1, seeds were asked to predict the average task completion time and quality rating after ChatGPT exposure in Noy and Zhang (2023). In round 2, seeds predicted how evaluators rated the novelty and usefulness of ChatGPT-assisted writing Hauser and Doshi (2024).

After eliciting prior beliefs, we revealed the treatment effects from Noy and Zhang (2023) or Hauser and Doshi (2024). Seeds then rated the evidence's credibility, interest, and value (for themselves and colleagues) on a 5-point Likert scale. We indicated that respondents would receive an infographic summarizing the study results along with the journal article upon survey completion. We then measured their intentions to read or share the materials and recommend ChatGPT to their colleagues. In round 2, additional questions explored when and why seeds might share the evidence. The survey concluded with demographic questions.

Shortly after the seed survey, we emailed all treated seeds links to the infographic and article PDF. We used a URL shortening service to monitor the number of clicks on each link over time, allowing us to track engagement levels on a daily basis.[12] The email was designed to be easily forwarded and included a note encouraging seeds to share it with colleagues.

Appendix Figure A.3 shows the distribution of our main outcomes.

Contact Survey Between two and three weeks after the seed survey, we invited each seed's colleagues within their division (*contacts*) to complete the "Contact Survey".

The contact survey assessed respondents' awareness of studies on the impact of ChatGPT on the productivity and creativity of skilled workers.[13] Specifically, we aimed to assess awareness of the study that had been disseminated to seeds in treated divisions. We manually coded contacts' answers provided in an open text box in response to the question "Are you familiar with any studies on the impact of generative AI tools such as ChatGPT on worker's productivity?" to determine which of the following study elements were correctly identified: names of the study's authors, institutional affiliations, the journal in which the article was published, and any noteworthy features of the study. This is a demanding test, since it requires sharing by seeds, recall of specific and detailed study information by contacts, and survey response of contacts. The second section of the survey explored respondents' current usage of and attitudes towards ChatGPT. The survey concluded by eliciting demographic information.

3 Main Results

3.1 Econometric Specification

In this section, we present the econometric specifications used to estimate the main treatment effects.

Treatment Variants To analyze outcomes specific to treated seeds who receive the evidence—such as seeds' intentions to share the evidence, their perceptions of the evidence, or actual engagement—we estimate the following specifications in the sample of treated divisions when pooling across both rounds of our experiment:

[12] A glitch in round 1 caused some follow-up emails to omit unique journal article links. As the issue was untraceable, we sent a follow-up email with the corrected link to all seeds who might have been affected, but decided to conservatively drop all engagement data for the article links from the analysis.

[13] Round 1 focused solely on productivity, while round 2 covered both productivity and creativity.

$$y_{id} = \alpha + \beta \text{senior treat}_d + X_d'\gamma + \varepsilon_{id}. \tag{1}$$

$$y_{id} = \alpha + \delta \text{info treat}_d + X_d'\gamma + \varepsilon_{id}. \tag{2}$$

where y_{id} is the outcome of interest; senior treat_d is an indicator for treated divisions in which we shared the evidence with a seed of senior rank; and info treat_d is an indicator for treated division in which we shared either positive information about ChatGPT adoption in the World Bank or additional information relating to the study's credibility[14]; finally, X_d represents a vector including fixed effects for the seed survey batches[15], and the strata fixed effects based on division size. Standard errors are clustered at the division level.

Treatment vs Control Our randomization strategy allows us to estimate overall rates of sharing by comparing contacts' recall of study features across treatment and control divisions using the following specification:

$$y_{id} = \alpha + \beta \text{treat}_d + X_d'\gamma + \varepsilon_{id}, \tag{3}$$

where treat_d is an indicator for any treated division.

Randomization Balance Appendix Tables A.1 and A.2 summarize balance checks for division and respondent covariates. Out of 56 comparisons, only 1 is significant at the 1 percent level and 2 at the 10 percent level, as would be expected given natural sampling variation.

3.2 Experimental Results

3.2.1 Predicted Treatment Effects

We conducted two expert surveys: one with World Bank researchers at an internal seminar, and another with social science researchers via the Social Science Prediction Platform (SSPP).[16]

The survey first outlines the experimental design and primary outcomes. Experts are shown the control group mean for each outcome—for example: "Among junior employees, 48% report being likely to share evidence on generative AI's impact." They then predict the corresponding outcome for the treatment group (e.g., senior seeds). Treatment order is randomized. Expert predictions are shown in Appendix Figure A.2.

[14] We estimate the latter specification using only round 1 data.

[15] The batch fixed effects account for potential differences between divisions where the surveyed seed responded immediately or after we had already invited one or two other employees to participate in the seed survey. These are defined separately for each round of the experiment. See Section 2.2 for further details.

[16] 69% of the SSPP sample are in academia (PhD students, postdocs, or tenure-track academics) and 88% identify as economists.

3.2.2 Demand Effects

To mitigate concerns about experimenter demand effects, we framed the seed survey as a general assessment of attitudes toward and use of ChatGPT at the World Bank, without revealing the study's objective. This approach proved effective: no seed correctly identified the purpose when asked directly.

3.2.3 Effects of Organizational Rank

First, we examine how seeding evidence with junior versus senior employees affects transmission. We begin by examining pooled effects across both experimental rounds. Panel A of Table 1 shows that senior seeds are 14.2 percentage points more likely ($p = 0.01$) to report being "likely" or "very likely" to share the evidence (Column 1). As shown in Appendix Figure A.2, experts underestimated the treatment effect of seeding a senior employee on sharing intentions, expecting only a 3 percentage point difference.

Consistent with higher sharing intentions, there is more actual engagement with the materials shared with senior seeds. Panel A shows a 15.8 percentage point increase ($p < 0.01$) in the likelihood that the infographic link was clicked more than once (Column 2) and 0.335 additional clicks ($p = 0.06$) on the link to the infographic (Column 3). Panels B and C confirm similar patterns across both rounds.

Next, we examine contacts' recall of key study features as a measure of information diffusion. Panel B of Table 2 shows the effect in round 1 is driven by divisions with a senior seed: contacts in these divisions are 5.0 percentage points more likely to recall at least one feature than those in treated divisions with a junior seed ($p = 0.03$). On the other hand, there is no recall observed in round 2. Expert predictions substantially exceed the actual effect in round 1, anticipating a 9.5 percentage point difference (Appendix Figure A.2). The treatment effect fades over time, with no significant differences in recall after four months (Appendix Table A.3).[17]

Mechanisms To better understand seeds' decisions to share, we asked them in round 2 to indicate their motivations for sharing and the barriers they faced. Appendix Figure A.4 shows that the most commonly cited reasons for sharing were the perceived value of the evidence to peers and to decision-makers. The most frequently mentioned barriers were the belief that colleagues lacked the time to engage with the evidence and a lack of time on the part of the seed themselves.

Senior seeds are significantly more likely than junior seeds to cite the value of the evidence to peers as a reason to share (Appendix Figure A.5, $p = 0.02$).[18] Appendix Figure A.5 also indicates that senior seeds are more likely to cite lack of time as a barrier to sharing ($p = 0.05$), but are suggestively less likely to mention their colleagues' time constraints ($p = 0.21$) or generating conflict among colleagues ($p = 0.03$). Together, this suggests senior seeds perceive the evidence to be more valuable to share and perceive fewer interpersonal frictions in doing so.

[17] We deviate from our pre-analysis plan in three ways (Appendix B.3): omitting seed-level controls to avoid bad controls, excluding engagement outcomes based on article clicks due to a technical glitch, and dichotomizing sharing intentions for precision.

[18] Consistent with this, Panel A of Table 3 shows that senior seeds evaluate the evidence more favorably across both rounds: all point estimates are positive, though not precisely estimated. Panels B and C indicate larger differences in round 1, when prior beliefs were more positive.

Appendix Figure A.6 shows senior and junior seeds also differ along several observable dimensions. Senior seeds are less likely to be female, more likely to hold a PhD, more likely to have managerial responsibilities, and more likely to report using ChatGPT in their work.[19] These differences in characteristics suggest that selection on observables may play a role in explaining estimated treatment effects. Appendix Figure A.7 suggests some of these characteristics are correlated with sharing behavior and engagement with study materials, possibly mediated by the treatment effects of seniority.

That said, these findings do not rule out the possibility that seniority itself has an independent effect on both the propensity and capacity to share information within organizations. Unobserved factors —such as higher status, greater respect or more central network positions —- may enhance senior employees' ability and inclination to disseminate evidence.

Indeed, senior seeds are more likely to hold formal managerial roles, which may lead them to view information dissemination as part of their responsibilities. To explore this channel, we look specifically at seeds whose position titles include any of the following words: "manager", "director", "chief", "head" and "team leader".[20] Appendix Figure A.8 provides suggestive evidence that seeds in managerial positions are more likely to report intentions to share the evidence; however, given the number of comparisons, this result should be interpreted with caution. However, Appendix Table A.4 confirms that the treatment effects persist when excluding seeds in managerial positions, suggesting senior seeds share evidence for reasons beyond their formal managerial duties. Senior seeds are also more likely to hold policy-oriented roles (e.g., "economist" or "specialist") focused on development policy rather than administrative or support functions such as legal or IT. Appendix Table A.6 shows treatment effects remain robust when restricting the sample to divisions where the seed held a policy-related position.

Adoption Panel A of Appendix Table A.8 shows that, when pooling responses across both rounds of the experiment, senior seeds are generally no more likely than junior seeds to report being "likely" or "very likely" to recommend ChatGPT based on the evidence provided. However, Panel B suggests that senior seeds in Round 1 are 10.5 percentage points more likely to express strong willingness to recommend ChatGPT ($p = 0.131$). Despite this suggestive difference in stated intentions, Columns 2 and 3 show no significant increase in actual adoption by contacts. That said, the estimates are imprecise, and we cannot rule out economically meaningful effects, suggesting that a subset of colleagues may have adopted ChatGPT in response to the treatment.[21]

3.2.4 Effects of positive versus negative information about adoption of ChatGPT

Next, we study the effects of altering seeds' beliefs about *(i)* colleagues' actual adoption of genAI and *(ii)* the extent to which their colleagues think World Bank employees should adopt ChatGPT at the workplace.

[19] Consistent with the gender differences documented in Humlum and Vestergaard (2025), we find women are less likely to report having a ChatGPT account.

[20] Seeds in these types of positions comprise 8 percent of our seed sample.

[21] Dar et al. (2022) also estimate imprecise average treatment effects on technology adoption, but document heterogeneous effects depending on perceived returns.

Seeds informed that ChatGPT adoption among colleagues is *low* may perceive evidence on the technology's benefits as socially less acceptable to share. The negative information treatment could then decrease their propensity to share the evidence. Conversely, seeds told adoption is *high* may view the same evidence as more valuable to share, anticipating that it aligns with prevailing norms and will be more positively received by colleagues. Consequently, these seeds may be more inclined to disseminate the evidence further.

Our information treatments successfully shifted seeds' incentivized beliefs about their colleagues' ChatGPT adoption and attitudes. Panel A of Table 3 presents treatment effects using data from both experimental rounds. Column 1 shows that seeds in the positive information condition estimated ChatGPT adoption among colleagues to be 12.3 percentage points higher ($p < 0.01$). Column 2 finds they predicted a 21.8 percentage point increase ($p < 0.01$) in colleagues supporting frequent ChatGPT use at the World Bank.

We also examine whether the information treatment influences seeds' second-order beliefs—specifically, their expectations about colleagues' priors on the effects of access to ChatGPT on productivity (task completion time and output quality) and creativity (novelty and usefulness), as estimated in Noy and Zhang (2023) and Hauser and Doshi (2024). In Panel B of Table 3, round 1 seeds in the positive information condition believe their colleagues estimate higher output quality (Column 5, $p < 0.01$), while the treatment has no clear effect on second-order beliefs about task completion time (Column 3). Similarly, Panel C of Table 3 shows that treated round 2 seeds believe their colleagues hold more positive views on the novelty and usefulness of a story written with access to ChatGPT (Columns 3 and 5; $p < 0.01$ for both). Consistent with this adjustment, seeds who receive more positive information also believe they personally hold more negative priors on the effects of ChatGPT relative to their colleagues (Columns 4 and 6).

We next examine the impact of the information treatment on seeds' sharing intentions and our measures of actual engagement with the study materials shared with seeds. Panel A of Table 1 shows that seeds in our positive information condition are no more likely to share the evidence than those receiving negative information. Specifically, we estimate a small, statistically not significant decrease of 1 percentage point in the likelihood of reporting a high intention to share the evidence (Column 1). Similarly, we observe a small and not significant effect on recorded link clicks (Columns 2 and 3), consistent across both experimental rounds (Panels B and C). Appendix Table A.9 shows no evidence of heterogeneity in the impact of the positive information treatment by seeds' seniority, although the estimates are somewhat imprecise. Table 2 further shows no treatment effect on contacts' recall of study features.

Appendix Figure A.2 highlights these null effects were not anticipated by our sample of experts. Experts expected the positive information treatment to increase seeds' intentions to disseminate the evidence by 9.7 percentage points, number of clicks on the infographic by 3.1 clicks, and the likelihood of recalling at least one study feature by 11.6 percentage points. We can reject the predicted impacts of all three outcomes ($p = 0.04$, $p < 0.01$ and $p < 0.01$, respectively).

3.2.5 Effects of Credibility

Next, we study the effects of sharing information designed to change seeds' perceptions of the credibility of the evidence. Panel B of Table 3 shows that treated seeds in round 1 rate the evidence's credibility 0.226 standard deviations higher (Column 7; $p = 0.11$) than those who receive no information about the article's publication status or other credibility signals. We interpret this finding as suggestive of a first-stage effect, implying that the treatment led to an increase in perceived credibility.

However, we find no significant effect on self-reported sharing intentions or actual engagement. Panel B of Table 1 estimates a small, statistically not significant increase of 1.2 percentage points in the likelihood of reporting a high intention to share the evidence (Column 1). Moreover, we find only small and imprecise effects on contacts' recall. Panel B of Table 2 shows that contacts in divisions where the seed received the credibility treatment are 1.6 percentage points more likely to recognize at least one notable study feature compared to those in control divisions (Column 2; $p = 0.45$). Consistent with the main findings, Appendix Table A.10 shows no evidence of heterogeneity in impact of the credibility treatment by seeds' seniority, although the estimates lack power to detect modest effect sizes.

While we estimate small positive effects on both sharing intentions and colleagues' recall, we observe a marginally significant negative effect on the likelihood that seeds recorded multiple clicks on the infographic link. One interpretation is that treated seeds, having received the credibility information, may have inferred that the study was already well-known within the organization — diminishing its perceived novelty and, in turn, reducing their motivation to engage further or share the evidence with others.

In sum, these findings suggest that shifting perceptions of the evidence's credibility does not play a first-order role in shaping seeds' decisions to share evidence with their colleagues. This stands in contrast to the emphasis on source credibility in the broader literature on information diffusion (e.g., Gentzkow and Shapiro, 2006) and, as shown in Appendix Figure A.2, was also not anticipated by our sample of experts. Experts expected the positive information treatment to increase seeds' intentions to disseminate the evidence by 8 percentage points, number of clicks on the infographic by 2.3 clicks, and the likelihood of recalling at least one study feature by 10.4 percentage points. We can reject the predicted values for the latter two outcomes. One possibility for the null effect is that credibility thresholds may already be saturated in our setting — respondents may have assumed that any evidence shared via a research survey had already been vetted and was therefore trustworthy.

3.2.6 Average Treatment Effects on Colleagues' Recall of Evidence

Across all divisions in round 1, contacts of treated seeds are suggestively more likely to recall notable features ($p = 0.21$; Panel A, Table 2). On average, 3% of contacts in treated divisions recall any feature of the study shared. Among those who do, roughly half recall more than one feature. The most frequently remembered elements are the authors' institutional affiliation and specific features of the study (Panel C, Figure A.3). Appendix Section B provides evidence that there are no statistically significant spillover effects from treated to untreated divisions.

4 Discussion

Our results underscore the central role of hierarchy in shaping informal information flows within organizations. Crucially, we examine both sharing intentions and actual diffusion: differences in colleagues' recall indicate that senior seeds were not only more likely to engage with and disseminate the evidence, but also more likely to be heard and remembered (Graeber et al., 2024). The absence of heterogeneous effects of the information treatments by seniority suggests that junior employees likely were not primarily constrained by conformity concerns or doubts about credibility. These findings have implications for organizational design and the diffusion of scientific evidence. While senior staff may be more effective messengers, our results point to potential bottlenecks when ideas originate lower in the hierarchy. Such frictions may be particularly costly in light of recent empirical findings that support standard hierarchical models in which agents are better informed than principals (Dal Bó et al., 2021).

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58.

_ , David Autor, Jonathon Hazell, and Pascual Restrepo, "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics*, 2022, *40* (S1), S293–S340.

Alatas, Vivi, Arun G Chandrasekhar, Markus Mobius, Benjamin A Olken, and Cindy Paladines, "Do celebrity endorsements matter? A Twitter experiment promoting vaccination in Indonesia," *The Economic Journal*, 2024, *134* (659), 913–933.

Angrist, Joshua D and Jörn-Steffen Pischke, *Mostly harmless econometrics: An empiricist's companion*, Princeton university press, 2009.

Atkin, D., A. Chaudhry, S. Chaudry, A. Khandelwal, and E. Verhoogen, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan," *The Quarterly Journal of Economics*, 2017, *132*, 1101—-1164.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "The diffusion of microfinance," *Science*, 2013, *341* (6144), 1236498.

_ , _ , _ , and _ , "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *The Review of Economic Studies*, 2019, *86* (6), 2453–2490.

Beaman, L., A. BenYishay, J. Magruder, and A. Mobarak, "Can Network Theory-Based Targeting Increase Technology Adoption?," *American Economic Review*, 2021, *111*, 1918–1943.

Beil, P. Rey, J. García-Hombrados, A. Martínez, B. Ozcan, T. Roldán, and M. Jansen, "Ideological Alignment and Evidence-Based Policy Adoption," *Working Paper*, 2025.

Beraja, Martin, Andrew Kao, David Y Yang, and Noam Yuchtman, "AI-tocracy," *The Quarterly Journal of Economics*, 2023, *138* (3), 1349–1402.

_ , David Y Yang, and Noam Yuchtman, "Data-intensive innovation and the state: Evidence from AI firms in China," *The Review of Economic Studies*, 2023, *90* (4), 1701–1723.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu, "Bureaucracy and development," *Annual Review of Economics*, 2022, *14* (1), 397–424.

Bick, Alexander, Adam Blandin, and David J Deming, "The rapid adoption of generative ai," Technical Report, National Bureau of Economic Research 2024.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, February 2013, *128* (1), 1–51.

Bó, Ernesto Dal, Frederico Finan, Nicholas Y Li, and Laura Schechter, "Information technology and government decentralization: Experimental evidence from Paraguay," *Econometrica*, 2021, *89* (2), 677–701.

Bottan, N. and R. Perez-Truglia, "Betting on the House: Subjective Expectations and Market Choices," *American Economic Journal: Applied Economics*, 2025, *17*, 459—500.

Breza, Emily, "Field experiments, social networks, and development," 2016.

_ and Arun G Chandrasekhar, "Social networks, reputation, and commitment: evidence from a savings monitors experiment," *Econometrica*, 2019, *87* (1), 175–216.

Buera, Francisco J. and Robert E. Lucas, "Idea Flows and Economic Growth," *Annual Review of Economics*, 2018, *10* (Volume 10, 2018), 315–345.

Bursztyn, Leonardo, Alessandra L González, and David Yanagizawa-Drott, "Misperceived social norms: Women working outside the home in Saudi Arabia," *American Economic Review*, 2020, *110* (10), 2997–3029.

_ and David Y Yang, "Misperceptions about others," *Annual Review of Economics*, 2022, *14* (1), 425–452.

Cai, Jing and Shing-Yi Wang, "Improving management through worker evaluations: Evidence from auto manufacturing," *The Quarterly Journal of Economics*, 2022, *137* (4), 2459–2497.

Conley, T. and C. Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 2010, *100*, 35–69.

Crafts, Nicholas, "Artificial intelligence as a general-purpose technology: an historical perspective," *Oxford Review of Economic Policy*, 2021, *37* (3), 521–536.

Crowley, D., J. Scott, E. Long, L. Green, A. Israel, L. Supplee, E. Jordan, K. Oliver, S. Guillot-Wright, B. Gay, R. Storace, N. Torres-Mackie, Y. Murphy, S. Donnay, J. Reardanz, R. Smith, K. McGuire, E. Baker, A. Antonopoulos, M. McCauley, and C. Giray, "Lawmakers' use of scientific evidence can be improved," *Proceedings of the National Academy of Sciences*, 2021, *118.*

Dar, Manzoor H, Alain de Janvry, Kyle Emerick, Erin M Kelley, and Elisabeth Sadoulet, "Casting a wider net: sharing information beyond social networks," Technical Report, CEGA Working Paper 2022.

DellaVigna, Stefano, Woojin Kim, and Elizabeth Linos, "Bottlenecks for Evidence Adoption," *Journal of Political Economy*, 2024, *132* (8).

Duflo, Esther, Daniel Keniston, Tavneet Suri, and Céline Zipfel, "Chat over coffee? diffusion of agronomic practices and market spillovers in Rwanda," Technical Report, National Bureau of Economic Research 2023.

Eil, David and Justin Rao, "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself," *American Economic Journal: Microeconomics*, 2011, *3*, 114–138.

et al. Ruggeri K., Stock F. Haslam S.A., "A synthesis of evidence for policy from behavioural science during COVID-19," *Nature*, 2024, *625*, 134—147.

Fong, Christian, "Expertise, networks, and interpersonal influence in congress," *The Journal of Politics*, 2020, *82* (1), 269–284.

Foster, A. and M. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 1995, *103*, 1176–1209.

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro et al., "Toward understanding the impact of artificial intelligence on labor," *Proceedings of the National Academy of Sciences*, 2019, *116* (14), 6531–6539.

Garicano, Luis and Luis Rayo, "Why organizations fail: Models and cases," *Journal of Economic Literature*, 2016, *54* (1), 137–192.

Gentzkow, Matthew and Jesse M Shapiro, "Media bias and reputation," *Journal of political Economy*, 2006, *114* (2), 280–316.

Goldfarb, Avi, Bledi Taska, and Florenta Teodoridis, "Artificial Intelligence in Health Care? Evidence from Online Job Postings," *AEA Papers and Proceedings*, May 2020, *110*, 400–404.

Graeber, Thomas, Shakked Noy, and Christopher Roth, "Lost in transmission," 2024.

Guillouet, Louise, Amit K Khandelwal, Rocco Macchiavello, Madhav Malhotra, and Matthieu Teachout, "Language barriers in multinationals and knowledge transfers," *Review of Economics and Statistics*, 2024, pp. 1–56.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart, "Designing information provision experiments," *Journal of economic literature*, 2023, *61* (1), 3–40.

Hardy, Morgan and Jamie McCasland, "It takes two: Experimental evidence on the determinants of technology diffusion," *Journal of Development Economics*, 2021, *149* (102600).

Hauser, Oliver and Anil Doshi, "Generative artificial intelligence enhances creativity but reduces the diversity of novel content," *Science Advances*, 2024, *10* (28), eadn5290.

Hjort, J., D. Moreira, G. Rao, and J.F. Santini, "How research affects policy: Experimental evidence from 2,150 Brazilian municipalities," *American Economic Review*, 2021, *111*, 1442–1480.

Humlum, Anders and Emilie Vestergaard, "The unequal adoption of ChatGPT exacerbates existing inequalities among workers," *Proceedings of the National Academy of Sciences*, 2025, *122* (1), e2414972121.

Juhasz, Reka, Mara P Squicciarini, and Nico Voigtlander, "Technology adoption and productivity growth: Evidence from industrialization in France," *Journal of Political Economy*, 2024, *132* (10), 3215–3259.

Kreitmeir, David H and Paul A Raschky, "The Unintended Consequences of Censoring Digital Technology–Evidence from Italy's ChatGPT Ban," *arXiv preprint arXiv:2304.09339*, 2023.

Levitt, Steven D. and Susanne Neckermann, "What field experiments have and have not taught us about managing workers," *Oxford Review of Economic Policy*, 04 2015, *30* (4), 639–657.

Macchiavello, Rocco and Ameet Morjaria, "The value of relationships: evidence from a supply shock to Kenyan rose exports," *American Economic Review*, 2015, *105* (9), 2911–2945.

Magruder, Jeremy R, "An assessment of experimental evidence on agricultural technology adoption in developing countries," *Annual Review of Resource Economics*, 2018, *10* (1), 299–316.

Mehmood, S., S. Naseer, and D. Chen, "AI Education as State Capacity: Experimental Evidence from Pakistan," 2024.

Menzel, Andreas, "Knowledge exchange and productivity spill-overs in Bangladeshi garment factories," *Journal of economic behavior & organization*, 2021, *185*, 721–746.

Mobarak, A. BenYishay A., "Social Learning and Incentives for Experimentation and Communication," *The Review of Economic Studies*, 2019, *86*, 976—-1009.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381*, 187–192.

Rao, Michelle, "Program Evaluations and Policy Spending," 2024.

Rogger, Daniel and Ravi Somani, "Hierarchy and Information," *Journal of Public Economics*, 2023, *219*, 104823.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, and Christopher T Stanton, "Workplace knowledge flows," *The Quarterly Journal of Economics*, 2020, *135* (3), 1635–1680.

Slavin, R., "Evidence-Based Education Policies: Transforming Educational Practice and Research," *Educational Researcher*, 2002, *31*, 15–21.

Thaler, M., M. Toma, and Victor Yaneng Wang, "Numbers Tell, Words Sell," 2025.

Vivalt, Eva and Aidan Coville, "How do policymakers update their beliefs?," *Journal of Development Economics*, 2023, *165*.

Tables

Table 1: Evidence Sharing by Seeded Employees

	Seed Survey	Infographic Engagement	
	Intend to share	Clicks > 1	Number of Clicks
	(1)	(2)	(3)
	Panel A: Pooled Treated Sample		
Senior Seed	0.142***	0.158***	0.335*
	(0.055)	(0.053)	(0.178)
Positive Info	-0.010	0.001	0.062
	(0.053)	(0.053)	(0.175)
Observations	336	335	335
Sample mean	0.55	0.39	1.53
	Panel B: Round 1 Treated Sample		
Senior Seed	0.156**	0.165**	0.353
	(0.070)	(0.071)	(0.274)
Positive Info	-0.029	-0.004	0.063
	(0.071)	(0.071)	(0.269)
Credibility Info	0.012	-0.126*	-0.347
	(0.071)	(0.071)	(0.256)
Observations	204	205	205
Sample mean	0.55	0.48	2.15
	Panel C: Round 2 Treated Sample		
Senior Seed	0.112	0.140*	0.301*
	(0.086)	(0.078)	(0.167)
Positive Info	0.001	-0.003	0.042
	(0.088)	(0.077)	(0.162)
Observations	132	130	130
Sample mean	0.55	0.24	0.56

Notes: This table reports OLS estimates of β and δ in equations (1) and (2). Each row corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic link. Panel A reports estimates for the pooled treated sample, while Panels B and C report estimates for rounds 1 and 2, respectively. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

Table 2: Information Diffusion to Seeded Employees' Colleagues

	Contact Survey	
Recognize Round 1 study:	At least 1 reported detail	Number of reported details
	(1)	(2)
	Panel A: Round 1 Treatment vs Control	
Treatment	0.020	-0.001
	(0.016)	(0.049)
Observations	380	380
Control mean	0.01	0.05
	Panel B: Round 1 Treated Sample	
Senior Seed	0.050**	0.078**
	(0.023)	(0.033)
Positive Info	0.020	0.024
	(0.022)	(0.031)
Credibility Info	0.016	0.016
	(0.021)	(0.031)
Observations	292	292
Sample mean	0.03	0.05

Notes: Panel A reports OLS estimates of β in equation (3) using the sample of contacts in the full sample. Panel B reports OLS estimates of β and δ in equations (1) and (2) using the sample of contacts in treated divisions. Each row in Panel B corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if at least 1 study feature (e.g., names of the study's authors, their institutional affiliations, the journal in which the article was published, and any noteworthy features of the study) was correctly reported by a contact; and in column 2, the number of study features correctly reported. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses.

Table 3: Information Treatment Impacts on Seeds' Second-Order Beliefs and Perceptions of Evidence

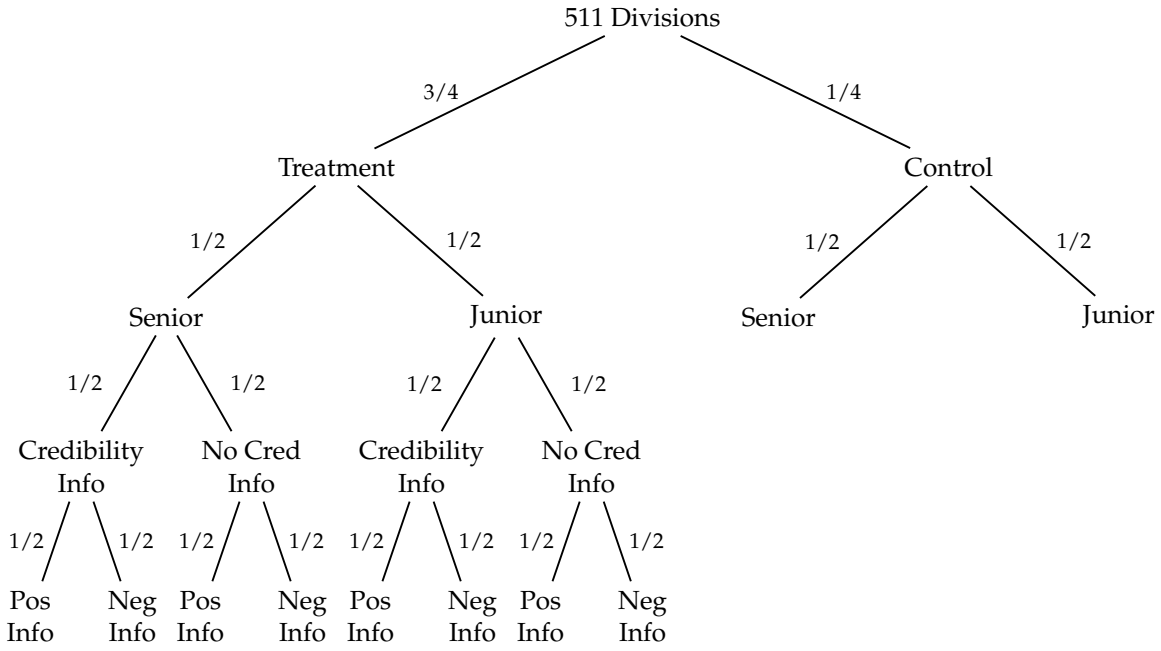
	Seed Survey									
	Beliefs about colleagues'...						*Beliefs about evidence*			
	ChatGPT Use	Support for ChatGPT Use	ChatGPT Prior 1		ChatGPT Prior 2		Credible	Interesting	Valuable to Colleagues	Valuable to Self
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Pooled Treated Sample										
Senior Seed	0.439	1.026					0.150	0.134	0.081	0.163
	(2.409)	(2.686)					(0.113)	(0.112)	(0.113)	(0.108)
Positive Info	12.286***	21.773***					-0.121	-0.000	-0.067	-0.101
	(2.538)	(2.492)					(0.109)	(0.112)	(0.110)	(0.110)
Observations	337	337					337	337	337	337
Sample mean	42.21	49.33								
Panel B: Round 1 Treated Sample										
			Task Completion Time		***Output Quality***					
			SO Belief	*Diff*	*SO Belief*	*Diff*				
Senior Seed	0.341	-0.138	-0.791	-0.066	0.003	-0.115	0.125	0.172	0.195	0.202
	(3.209)	(3.454)	(1.296)	(1.021)	(0.135)	(0.127)	(0.148)	(0.145)	(0.144)	(0.141)
Positive Info	15.510***	25.418***	-1.096	-0.606	0.444***	-0.351***	-0.074	0.098	0.107	-0.031
	(3.106)	(3.020)	(1.306)	(0.980)	(0.127)	(0.124)	(0.140)	(0.140)	(0.141)	(0.140)
Credibility Info	1.191	-0.074	1.119	-0.854	-0.155	0.178	0.226	0.211	0.083	0.117
	(3.188)	(3.503)	(1.270)	(0.989)	(0.132)	(0.130)	(0.141)	(0.142)	(0.142)	(0.139)
Observations	205	205	205	205	205	205	205	205	205	205
Sample mean	39.5	48.03	18.97	-0.61	4.35	-0.10				
Panel C: Round 2 Treated Sample										
			Novelty		***Usefulness***					
			SO Belief	*Diff*	*SO Belief*	*Diff*				
Senior Seed	0.099	1.629	0.265	-0.395*	0.861***	-0.442*	0.156	0.064	-0.058	0.138
	(4.127)	(4.190)	(0.304)	(0.233)	(0.272)	(0.229)	(0.182)	(0.174)	(0.178)	(0.174)
Positive Info	6.433	15.182***	0.960***	-0.459**	0.674**	-0.514**	-0.220	-0.136	-0.313*	-0.199
	(4.252)	(4.004)	(0.300)	(0.227)	(0.261)	(0.206)	(0.178)	(0.170)	(0.175)	(0.168)
Observations	132	132	132	132	132	132	132	132	132	132
Sample mean	46.41	51.34	4.05	0.07	4.97	0.06				

Notes: This table reports OLS estimates of β and δ in equations (1) and (2). Each row corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. In Panels A through C, the dependent variable is: in column 1, seed employees' beliefs about the % of colleagues who use ChatGPT for their work; in column 2, seed employees' beliefs about the % of colleagues who somewhat or strongly agree ChatGPT should be used often in their work; in columns 7 to 10, a standardized z-score measured on a 1-5 scale of seeded employees' beliefs the evidence is: *credible*, *interesting*, *valuable to colleagues*, and *valuable to one's self*. In Panel B, the dependent variable is: in column 3, seed employees' beliefs about colleagues' priors on the effects of access to ChatGPT on task completion time; in column 4, the difference between seed employees' own and colleagues' priors on these effects; in column 5, seed employees' beliefs about colleagues' priors on the effects of access to ChatGPT on output quality; and in column 6, the difference between seed employees' own and colleagues' priors on these effects. In Panel C, the dependent variable is: in column 3, seed employees' beliefs about colleagues' priors on the effects of access to ChatGPT on the novelty of a story; in column 4, the difference between seed employees' own and colleagues' priors on these effects; in column 5, seed employees' beliefs about colleagues' priors on the effects of access to ChatGPT on usefulness; and in column 6, the difference between seed employees' own and colleagues' priors on these effects. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

A Additional Exhibits

Figures

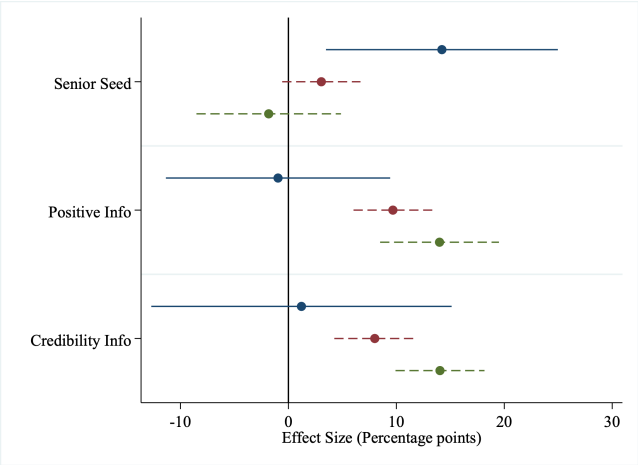
Figure A.1: Experimental design



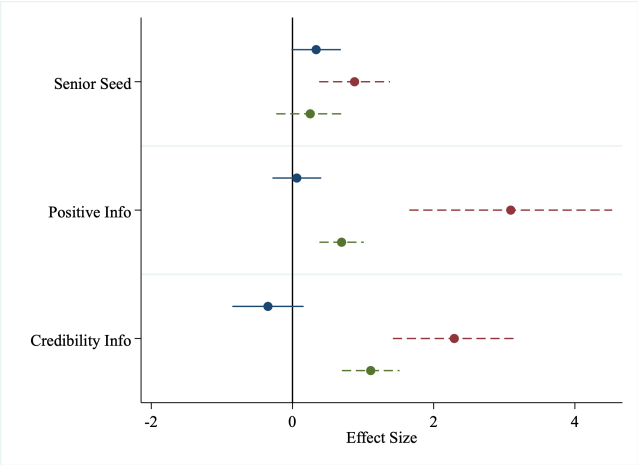
Notes: This figure presents the experimental design for Round 1. Randomization is at the level of World Bank division. Divisions are first assigned to a treatment (75%) or control group (25%). A seeded employee in a treated division receives scientific evidence about the impacts of genAI in the workplace in a survey, while a seeded employee in a control division receives only the survey. In 50% of divisions in the treated and control groups, the seeded employee is senior, while in the remaining 50%, the seeded employee is junior. We also implemented two cross-randomized treatments: In the first cross-randomized treatment, we randomly varied where we shared more positive or more negative information on colleagues' adoption of and views on generative AI. In the second cross-randomized treatment, we randomly varied whether we shared information regarding the credibility of the evidence. In round 2 (not displayed here), we re-randomized the organizational rank and information conditions within the treatment group, while excluding the credibility condition. We do not vary the credibility of the study in round 2 as Hauser and Doshi (2024) was not published at the time of the experiment. See Section 2.4 for details.

Figure A.2: Expert predictions vs Estimated treatment effects

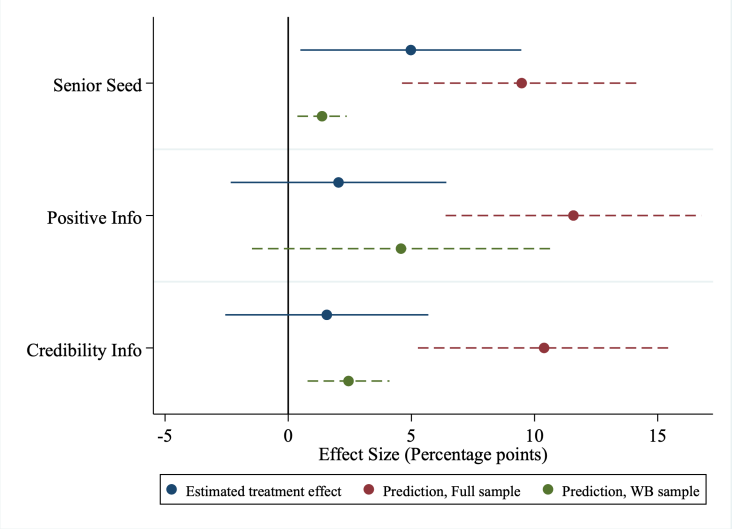
(a) Seeded employees' intention to share



(b) Number of clicks on infographic



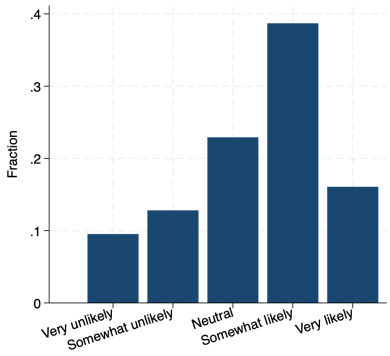
(c) At least 1 study detail reported by contacts



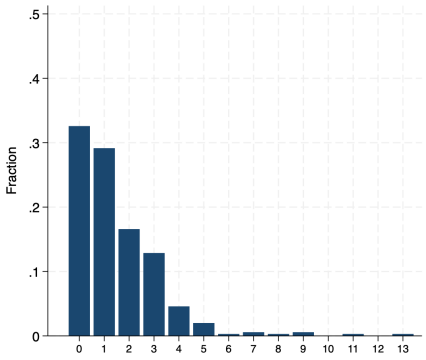
Notes: This figure presents treatment effects as estimated in Tables 1 and 2 for three outcomes and the corresponding expert predictions from the Social Science Prediction Platform and World Bank Surveys. Treatment effects and 95% confidence intervals are displayed in navy. Expert predictions from the full sample of 72 Social Science Prediction Platform and World Bank respondents are displayed in maroon, while expert predictions from the World Bank sample of 23 respondents are displayed in green. In Panel (a) the outcome is a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in Panel (b), the outcome is the number of clicks on the infographic link; and in Panel (c), the outcome is a dummy variable for at least 1 study detail reported correctly by contacts. Expert predictions for number of clicks on the infographic are winsorized at the 90% level to address extreme outliers.

Figure A.3: Measures of information transmission and diffusion

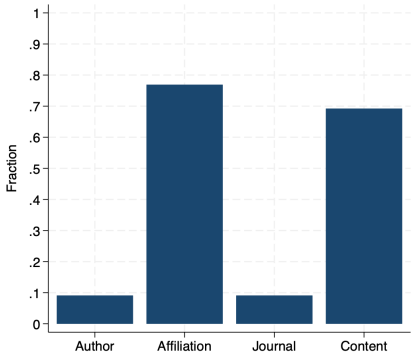
(a) Seeded employees' intention to share



(b) Number of clicks on infographic

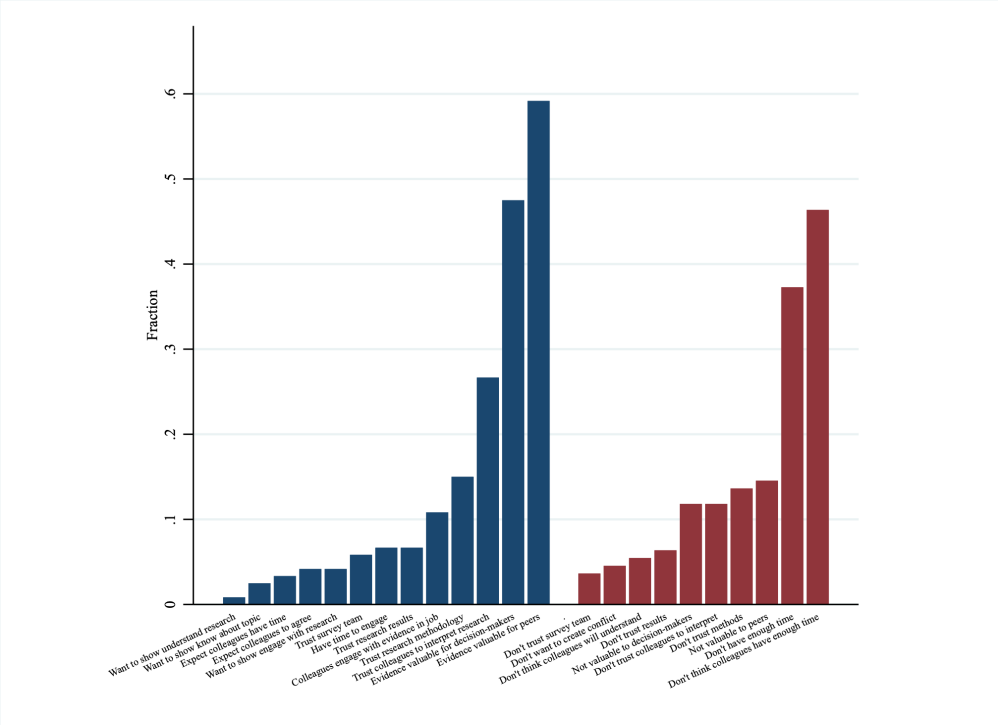


(c) Number of study details reported by contacts



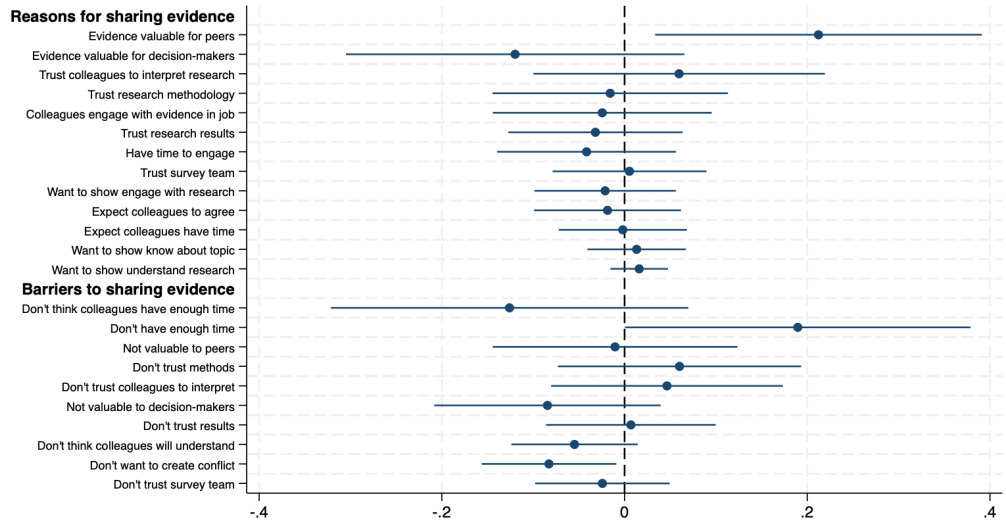
Notes: This figure presents the distributions of different measures of information transmission and diffusion. The measure is: in Panel (a), a seeded employee's self-reported intention to share the evidence on genAI's impact measured on a 5-point Likert scale; in Panel (b), the number of clicks on the infographic link; and in Panel (c), the number of study features (i.e., names of the study's authors, their institutional affiliations, the journal in which the article was published, or any noteworthy features of the study was correctly reported) correctly reported by contacts. The sample is seeded employees in treated divisions in Panel A; infographic links in treated divisions in Panel B; and contacts who report at least 1 study detail correctly in Panel C. See Section 3 for details.

Figure A.4: Reasons for and barriers to information transmission



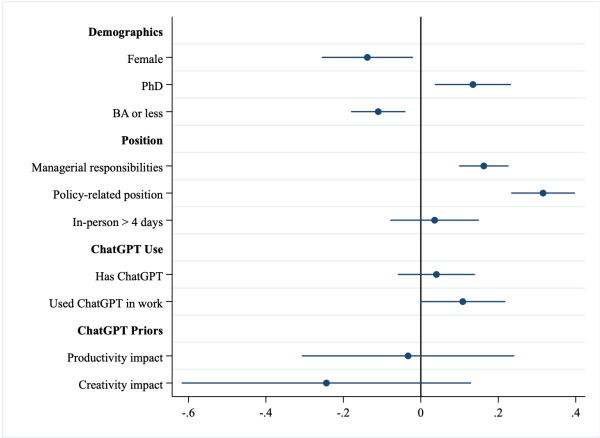
Notes: This figure presents seeded employees' self-reported reasons (navy) and barriers (maroon) to sharing genAI evidence with colleagues. The sample is seeded employees in treated divisions in round 2; these questions were not asked in round 1. Respondents could select multiple answers. The most common reason is the evidence's value to peers; the most common barrier is colleagues' time constraints.

Figure A.5: Senior and junior seeds suggestively identify different reasons for & barriers to information transmission (Correlations)



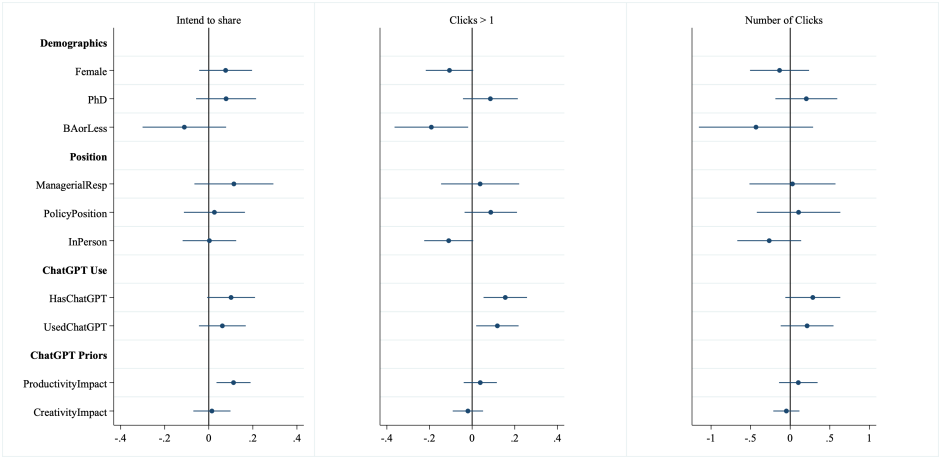
Notes: The figure presents correlations between senior rank and perceived reasons/barriers to information transmission. Each row shows a point estimate from a separate regression with a 95% confidence interval. Senior seed employees are more likely than junior ones to cite "Evidence valuable for peers" as a reason to share. They are also more likely to cite time constraints and less likely to cite fear of conflict as barriers.

Figure A.6: Senior seeds differ from junior seeds on observables (Correlations)



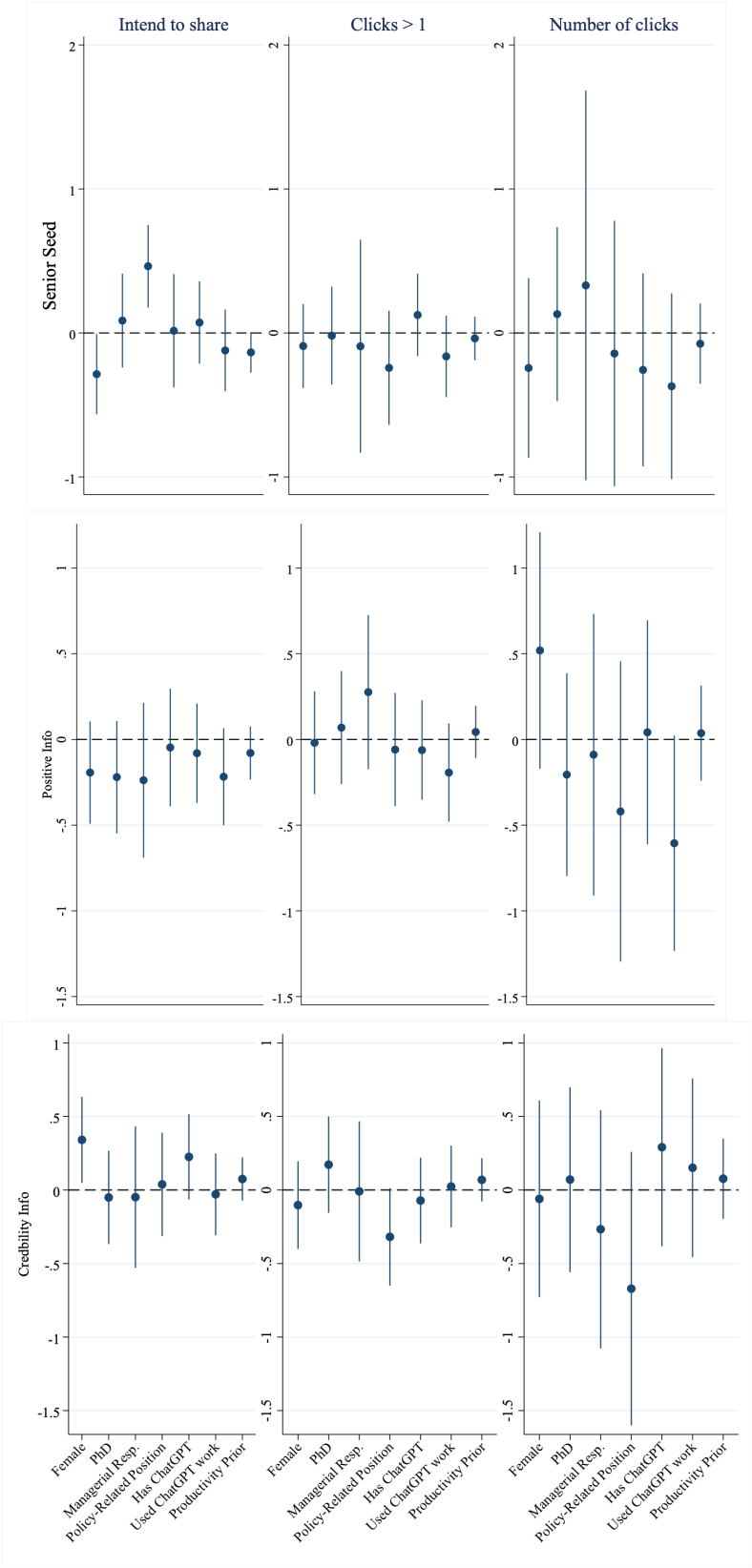
Notes: The figure presents correlations between senior rank and observable characteristics. Each row displays the point estimate from a separate regression alongside the 95% confidence interval.

Figure A.7: Seed observables, genAI priors, and information transmission (Correlations)



Notes: The figure presents correlations between seed observables and genAI priors with measures of information transmission. Each row displays the point estimate from a separate regression alongside the 95% confidence interval. The measures of information transmission are: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic.

Figure A.8: Treatment heterogeneity by seed characteristics is minimal



Notes: These figures show estimated treatment effects by seeded employees' characteristics in round 1 alongside the 95% confidence interval. The displayed treatment effect is the interaction between the seed's characteristic and the treatment. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, a standardized measure of number of clicks on the infographic link. Seed observables are defined as in Figure A.6.

Tables

Table A.1: Randomization Balance, Population

	Num. Employees in Division	Senior Employee	Managerial Responsibilities	Policy-Related Position
	(1)	(2)	(3)	(4)
		Panel A: Treatment vs Control		
Treatment	0.114	-0.029	0.025***	-0.029
	(0.938)	(0.023)	(0.008)	(0.043)
Observations	511	6254	6254	6254
Sample mean	12.24	0.51	0.08	0.74
		Panel B: Round 1 Treated Sample		
Senior Seed	0.610	0.040	0.003	-0.039
	(1.132)	(0.025)	(0.013)	(0.050)
Positive Info	-0.594	-0.034	-0.007	-0.050
	(1.130)	(0.025)	(0.013)	(0.050)
Credibility Info	0.589	-0.038	-0.007	-0.049
	(1.131)	(0.024)	(0.013)	(0.049)
Observations	381	4674	4674	4674
Sample mean	12.27	0.50	0.08	0.74
		Panel C: Round 2 Treated Sample		
Senior Seed	-0.724	-0.002	0.010	0.007
	(1.057)	(0.026)	(0.015)	(0.053)
Positive Info	-1.230	-0.015	-0.015	-0.015
	(1.056)	(0.026)	(0.014)	(0.052)
Observations	348	3906	3906	3906
Sample mean	11.22	0.50	0.08	0.73

Notes: Panel A reports balance checks for all 511 divisions in sample assigned to treatment or control, while Panels B and C report balance checks for the 381 and 348 divisions in the rounds 1 and 2 treated samples, respectively. The sample is at the division-level in column 1 and at the individual-level in columns 2 to 4. Regressions in columns 2 to 4 include strata fixed effects based on division size. The dependent variable is: in column 1, the total number of employees in a division; in column 2, a dummy variable equal to 1 if the employee is senior; in column 3, a dummy variable equal to 1 if the employee has managerial responsibilities; and in column 4, a dummy variable equal to 1 if the employee holds a policy-related position. Managerial responsibilities are identified by the words "manager", "director", "chief", "head" or "team leader" listed in the position title. Policy-related positions exclude employees whose roles are associated with legal, information technology, accounting, health, or counseling functions. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Robust standard errors in parentheses in column 1. Robust standard errors clustered by division in parentheses in columns 2 to 4.

Table A.2: Randomization Balance, Seed Sample

	Female	PhD	BA or Less Educ.	Managerial Resp.	Policy-Related Position	In Person $\geq$ 4 days	Has ChatGPT	Used ChatGPT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Pooled Treated Sample				
Positive Info	0.033	-0.020	-0.007	-0.016	-0.027	-0.015	-0.047	-0.091*
	(0.057)	(0.054)	(0.035)	(0.030)	(0.043)	(0.056)	(0.055)	(0.053)
Observations	295	298	298	337	337	301	337	337
Sample mean	0.57	0.26	0.11	0.09	0.78	0.63	0.64	0.52
				Panel B: Round 1 Treated Sample				
Positive Info	0.025	-0.041	0.007	-0.008	-0.050	-0.043	-0.079	-0.120*
	(0.071)	(0.063)	(0.046)	(0.042)	(0.058)	(0.067)	(0.069)	(0.071)
Credibility Info	0.057	0.022	0.008	-0.049	-0.048	-0.048	-0.093	0.015
	(0.072)	(0.062)	(0.043)	(0.042)	(0.057)	(0.067)	(0.068)	(0.071)
Observations	199	201	201	205	205	204	205	205
Sample mean	0.59	0.26	0.11	0.10	0.78	0.64	0.63	0.52
				Panel C: Round 2 Treated Sample				
Positive Info	0.061	0.009	-0.034	-0.036	0.013	0.037	0.016	-0.041
	(0.106)	(0.095)	(0.063)	(0.049)	(0.072)	(0.105)	(0.084)	(0.088)
Observations	96	97	97	132	132	97	132	132
Sample mean	0.51	0.27	0.09	0.08	0.80	0.60	0.66	0.53

Notes: Panel A reports balance checks for all seeds who completed the survey and provided demographic information, while Panels B and C reports balance checks for rounds 1 and 2, respectively. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The dependent variable is: in column 1, a dummy equal to 1 if the respondent is female; in column 2, a dummy equal to 1 if the respondent has a PhD; in column 3, a dummy equal to 1 if the respondent has a bachelor's degree or lower; in column 4, a dummy equal to 1 if the respondent has managerial responsibilities; in column 5, a dummy equal to 1 if the respondent holds a policy-related position; in column 6, a dummy equal to 1 if the respondent reports working in the office for four or more days per week; in column 7, a dummy equal to 1 if the respondent has a ChatGPT account; and in column 8, a dummy eqal to 1 if the respondent reports having used ChatGPT at work. Managerial responsibilities are identified by the words "manager", "director", "chief", "head" or "team leader" listed in the position title. Policy-related positions exclude employees whose roles are associated with legal, information technology, accounting, health, or counseling functions. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

Table A.3: Information Diffusion to Seeded Employees' Colleagues, Long-Run Effects

	Contact Survey	
Recognize Round 1 study:	At least 1 reported detail	Number of reported details
	(1)	(2)
	Panel A: Round 1 Treatment vs Control	
Treatment	-0.012	-0.007
	(0.017)	(0.019)
Observations	204	204
Control mean	0.02	0.02
	Panel B: Round 1 Treated Sample	
Senior Seed	0.012	0.025
	(0.013)	(0.026)
Positive Info	-0.014	-0.027
	(0.014)	(0.028)
Credibility Info	0.016	0.032
	(0.017)	(0.033)
Observations	153	153
Control mean	0.01	0.01

Notes: Panel A reports OLS estimates of β in equation (3) using the sample of contacts in the full sample. Panel B reports OLS estimates of β and δ in equations (1) and (2) using the sample of contacts in treated divisions. Each row in Panel B corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The sample is restricted to contacts who were surveyed in round 2 and who are colleagues of employees seeded in round 1. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if at least 1 study feature (e.g., names of the study's authors, their institutional affiliations, the journal in which the article was published, and any noteworthy features of the study) was correctly reported by a contact; and in column 2, the number of study features correctly reported. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses.

Table A.4: Senior seeded employees without managerial responsibilities share evidence

	Seed Survey	**Infographic Engagement**	
	Intend to share	Clicks > 1	Number of Clicks
	(1)	(2)	(3)
	Panel A: Pooled Treated Sample		
Senior Seed	0.116**	0.166***	0.340*
	(0.059)	(0.056)	(0.189)
Positive Info	0.004	-0.023	0.056
	(0.056)	(0.056)	(0.188)
Observations	304	303	303
Sample mean	0.54	0.38	1.52
	Panel B: Round 1 Treated Sample		
Senior Seed	0.151**	0.167**	0.334
	(0.074)	(0.075)	(0.295)
Positive Info	-0.008	-0.033	0.070
	(0.075)	(0.075)	(0.288)
Credibility Info	0.019	-0.120	-0.289
	(0.075)	(0.075)	(0.279)
Observations	183	184	184
Sample mean	0.55	0.47	2.13
	Panel C: Round 2 Treated Sample		
Senior Seed	0.058	0.161*	0.340*
	(0.092)	(0.084)	(0.181)
Positive Info	0.006	-0.012	0.031
	(0.094)	(0.081)	(0.173)
Observations	121	119	119
Sample mean	0.51	0.24	0.57

Notes: This table reports OLS estimates of β and δ in equations (1) and (2). Each row corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The sample excludes employees with managerial responsibilities. Managerial responsibilities are identified by the words "manager", "director", "chief", "head" or "team leader" listed in the position title. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic link. Panel A reports estimates for the pooled treated sample, while Panels B and C report estimates for Rounds 1 and 2, respectively. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

Table A.5: Information Diffusion to Seeded Employees' Colleagues, Exclude Seeds with Managerial Responsibilities

	Contact Survey	
Recognize Round 1 study:	At least 1 reported detail	Number of reported details
	(1)	(2)
	Panel A: Round 1 Treatment vs Control	
Treatment	0.021	0.001
	(0.016)	(0.050)
Observations	342	342
Control mean	0.01	0.05
	Panel B: Round 1 Treated Sample	
Senior Seed	0.054**	0.086**
	(0.025)	(0.037)
Positive Info	0.015	0.019
	(0.024)	(0.034)
Credibility Info	0.010	0.010
	(0.022)	(0.034)
Observations	255	255
Control mean	0.04	0.05

Notes: Panel A reports OLS estimates of β in equation (3) using the sample of contacts in the full sample. Panel B reports OLS estimates of β and δ in equations (1) and (2) using the sample of contacts in treated divisions. Each row in Panel B corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The sample excludes divisions where the seed has managerial responsibilities. Managerial responsibilities are identified by the words "manager", "director", "chief", "head" or "team leader" listed in the position title. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if at least 1 study feature (e.g., names of the study's authors, their institutional affiliations, the journal in which the article was published, and any noteworthy features of the study) was correctly reported by a contact; and in column 2, the number of study features correctly reported. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses.

Table A.6: Senior seeded employees in policy-related positions share evidence

	Seed Survey	Infographic Engagement	
	Intend to share	Clicks > 1	Number of Clicks
	(1)	(2)	(3)
	Panel A: Pooled Treated Sample		
Senior Seed	0.162**	0.134**	0.324*
	(0.064)	(0.062)	(0.173)
Positive Info	-0.005	-0.005	-0.026
	(0.061)	(0.062)	(0.179)
Observations	264	263	263
Sample mean	0.55	0.40	1.54
	Panel B: Round 1 Treated Sample		
Senior Seed	0.176**	0.132	0.338
	(0.082)	(0.083)	(0.256)
Positive Info	-0.043	-0.016	-0.100
	(0.080)	(0.081)	(0.265)
Credibility Info	0.019	-0.189**	-0.547**
	(0.080)	(0.081)	(0.236)
Observations	159	159	159
Sample mean	0.55	0.50	2.14
	Panel C: Round 2 Treated Sample		
Senior Seed	0.114	0.125	0.266
	(0.100)	(0.091)	(0.195)
Positive Info	0.015	0.001	0.081
	(0.097)	(0.087)	(0.184)
Observations	105	104	104
Sample mean	0.55	0.26	0.62

Notes: This table reports OLS estimates of β and δ in equations (1) and (2). Each row corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The sample restricts to employees with policy-related positions. Policy-related positions exclude employees whose roles are associated with legal, information technology, accounting, health, or counseling functions. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic link. Panel A reports estimates for the pooled treated sample, while Panels B and C report estimates for Rounds 1 and 2, respectively. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

Table A.7: Information Diffusion to Seeded Employees' Colleagues, Seeds in Policy-Related Positions

	Contact Survey	
Recognize Round 1 study:	At least 1 reported detail	Number of reported details
	(1)	(2)
	Panel A: Round 1 Treatment vs Control	
Treatment	0.019	-0.009
	(0.018)	(0.058)
Observations	312	312
Control mean	0.01	0.05
	Panel B: Round 1 Treated Sample	
Senior Seed	0.051**	0.081**
	(0.024)	(0.037)
Positive Info	0.022	0.025
	(0.027)	(0.037)
Credibility Info	0.025	0.028
	(0.024)	(0.035)
Observations	239	239
Control mean	0.04	0.06

Notes: Panel A reports OLS estimates of β in equation (3) using the sample of contacts in the full sample. Panel B reports OLS estimates of β and δ in equations (1) and (2) using the sample of contacts in treated divisions. Each row in Panel B corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. The sample restricts to divisions where the seed holds a policy-related position. Policy-related positions exclude employees whose roles are associated with legal, information technology, accounting, health, or counseling functions. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if at least 1 study feature (e.g., names of the study's authors, their institutional affiliations, the journal in which the article was published, and any noteworthy features of the study) was correctly reported by a contact; and in column 2, the number of study features correctly reported. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses.

Table A.8: ChatGPT Adoption

	Seed Survey	Contact Survey	
	Recommends ChatGPT	Has ChatGPT	Used ChatGPT
	(1)	(2)	(3)
	Panel A: Pooled Treated Sample		
Senior Seed	0.037	-0.025	0.010
	(0.054)	(0.048)	(0.054)
Positive Info	-0.056	-0.016	-0.025
	(0.053)	(0.048)	(0.054)
Observations	336	477	477
Sample mean	0.61	0.66	0.50
	Panel B: Round 1 Treated Sample		
Senior Seed	0.105	0.014	0.042
	(0.070)	(0.068)	(0.072)
Positive Info	-0.024	0.056	0.095
	(0.069)	(0.060)	(0.068)
Credibility Info	0.040	0.048	0.023
	(0.069)	(0.067)	(0.073)
Observations	204	284	284
Sample mean	0.62	0.65	0.50
	Panel C: Round 2 Treated Sample		
Senior Seed	-0.047	-0.091	-0.050
	(0.087)	(0.074)	(0.074)
Positive Info	-0.104	-0.113	-0.201***
	(0.087)	(0.075)	(0.070)
Observations	132	193	193
Sample mean	0.61	0.68	0.49

Notes: This table reports OLS estimates of β and δ in equations (1) and (2) using the sample of seeds in treated divisions in column 1 and the sample of contacts in treated divisions in column 2. Each row corresponds to a separate regression. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. *Senior Seed* estimates the effect of sharing scientific evidence on the impacts of generative AI with a senior employee. *Positive info* estimates the effect of receiving positive information about colleagues' adoption and beliefs about ChatGPT at the workplace. *Credibility info* estimates the effect of receiving information about the credibility of the evidence. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "somewhat likely" or "very likely" to recommend ChatGPT to their colleagues; in column 2, a dummy variable equal to 1 if the contact reports having a ChatGPT account; and in column 3, a dummy variable equal to 1 if the contact reports having used ChatGPT at work. Panel A reports estimates for the pooled treated sample, while Panels B and C report estimates for rounds 1 and 2, respectively. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Column 1 of Panel A and Column 2 of Panels A, B, and C. Robust standard errors in parentheses in Columns 2-3 of Panel A.

Table A.9: Positive information does not increase junior seeds' sharing

	Seed Survey	**Infographic Engagement**	
	Intend to share	Clicks > 1	Number of Clicks
	(1)	(2)	(3)
	Panel A: Pooled Treated Sample		
Positive Info	-0.024	-0.049	0.125
	(0.079)	(0.071)	(0.268)
Senior Seed x Positive Info	0.026	0.094	-0.124
	(0.110)	(0.102)	(0.354)
Senior Seed	0.130*	0.113	0.394*
	(0.073)	(0.073)	(0.225)
Observations	336	335	335
Sample mean	0.55	0.39	1.53
	Panel B: Round 1 Treated Sample		
Positive Info	-0.018	-0.038	0.217
	(0.104)	(0.100)	(0.426)
Senior Seed x Positive Info	-0.028	0.058	-0.314
	(0.143)	(0.143)	(0.549)
Senior Seed	0.169*	0.138	0.498
	(0.096)	(0.099)	(0.329)
Observations	204	205	205
Sample mean	0.55	0.48	2.15
	Panel C: Round 2 Treated Sample		
Positive Info	-0.050	-0.071	-0.029
	(0.130)	(0.098)	(0.206)
Senior Seed x Positive Info	0.107	0.144	0.163
	(0.178)	(0.153)	(0.327)
Senior Seed	0.064	0.073	0.228
	(0.119)	(0.105)	(0.227)
Observations	132	130	130
Sample mean	0.55	0.24	0.56

Notes: This table reports OLS estimates of the organizational rank (*senior seed*) and positive information about ChatGPT adoption (*positive info*) treatments interacted. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic link. Panel A reports estimates for the pooled treated sample, while Panels B and C report estimates for rounds 1 and 2, respectively. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

Table A.10: Credibility information does not increase junior seeds' sharing

	Seed Survey	Infographic Engagement	
	Intend to share	Clicks > 1	Number of Clicks
	(1)	(2)	(3)
	Round 1 Treated Sample		
Credibility Info	-0.082	-0.085	0.077
	(0.105)	(0.103)	(0.420)
Senior Seed x Credibility Info	0.200	-0.058	-0.765
	(0.145)	(0.144)	(0.534)
Senior Seed	0.053	0.187*	0.727*
	(0.105)	(0.104)	(0.385)
Observations	204	205	205
Sample mean	0.55	0.48	2.15

Notes: This table reports OLS estimates of the organizational rank (*senior seed*) and credibility information (*crediblity info*) treatments interacted. The dependent variable is: in column 1, a dummy variable equal to 1 if the seeded employee reports being "likely" or "very likely" to share the evidence; in column 2, a dummy variable equal to 1 if the infographic link was clicked more than once; and in column 3, the number of clicks on the infographic link. Estimates are for round 1. Regressions include strata fixed effects based on division size, and fixed effects for seed survey batches. See Section 3 for details. * p<0.1, ** p<0.05, *** p<0.01. Robust standard errors clustered by division in parentheses in Panel A. Robust standard errors in parentheses in Panels B and C.

B Additional Results

B.1 Descriptive analysis of ChatGPT use

This section presents descriptive statistics on ChatGPT use based on the combined seed and contact survey samples collected between late 2023 and early 2024 ($N = 1,335$).

Overall, 63% of respondents report having a ChatGPT account. Among these, approximately one-fourth hold a paid subscription, while the remainder use the free version. Close to half of all respondents indicate having used generative AI in their professional work. However, adoption rates vary substantially both across Program Management Units (PMUs) and by individual characteristics. For instance, among PMUs with at least 25 employees participating in the survey, reported work-related ChatGPT use ranges from 36% to 63%.

Appendix Figure B.1 presents individual-level correlates of ChatGPT use at work for the sample of seeds in treated divisions. Those who have used ChatGPT are more likely to hold a PhD, more likely to hold a senior position, slightly more likely to occupy a policy-related position, and tend to hold more positive priors regarding the effects of ChatGPT on productivity and creativity. In contrast, usage rates do not differ systematically between male and female staff.

59% of respondents report that they "somewhat agree" or "strongly agree" that generative AI tools such as ChatGPT should be used in their work. Support for ChatGPT use is higher among men, while views are broadly similar across seniority levels as shown in Appendix Figure B.2.

On a scale from -10 (extremely negative) to 10 (extremely positive), respondents rate the expected impact of ChatGPT on skilled worker productivity to be an average of 4.84, and on skilled worker cre-

Figure B.1: Survey respondent characteristics and ChatGPT use (Correlations)

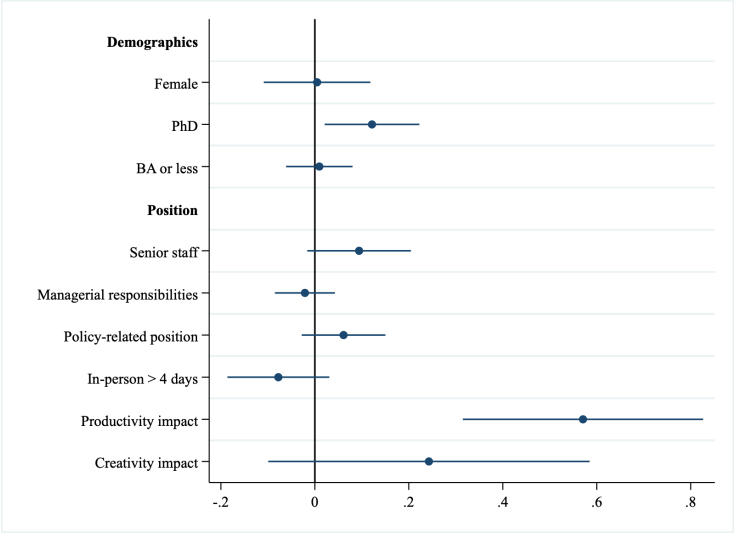
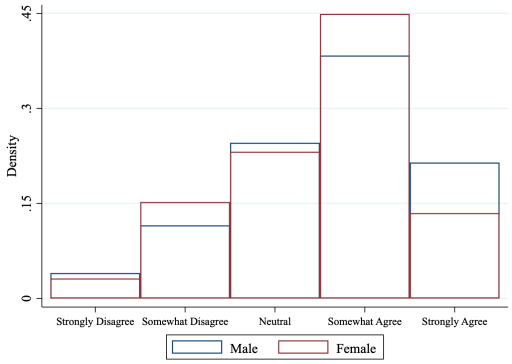
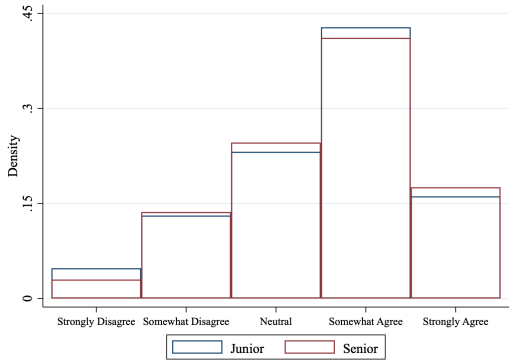


Figure B.2: Support for ChatGPT use at work

(a) By Gender



(b) By Rank



ativity to be 2.47. The full distributions are presented in Appendix Figure B.3. ChatGPT priors exhibit no meaningful correlation with observable individual-level demographics or position attributes.

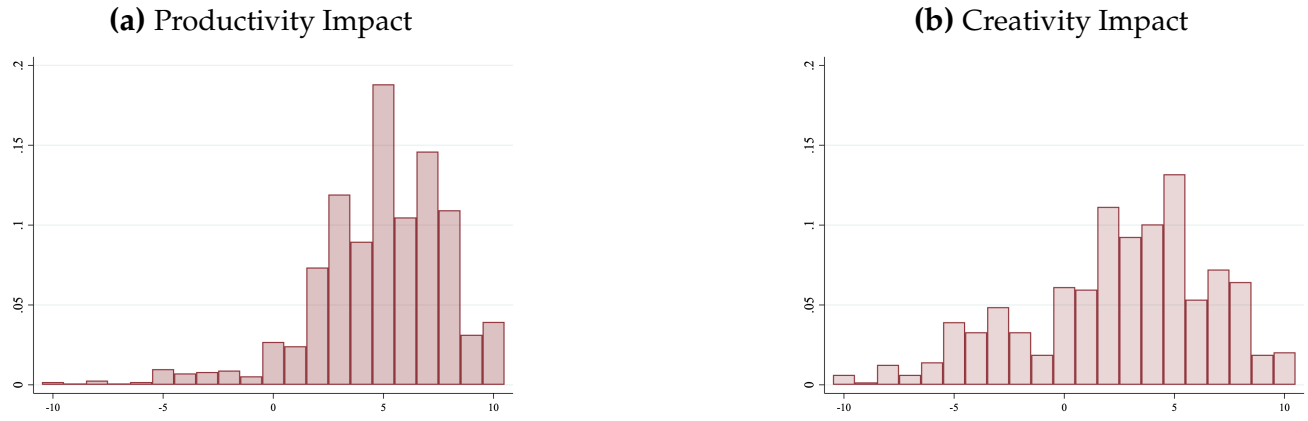
B.2 Spillover effects

Our randomization varied the share of treated divisions within Program Manageument Units (PMUs), enabling us to assess cross-division spillovers. We estimate:

$$y_{id} = \alpha + \beta_1 \text{treated}_d + \beta_2 \text{saturation}_d + \beta_3 \text{treated}_d \times \text{saturation}_d + \delta_d + \varepsilon_{id}, \tag{B.1}$$

where saturation_d is either the share of divisions within a PMU assigned to treatment or the share of divisions within a PMU actually treated. We estimate these specifications using the entire sample of contacts from round 1, including those who completed the survey in round 1, but belonged to divisions where no seed completed the survey. These contacts are assigned to the control group.

Figure B.3: Priors on impacts of ChatGPT



Appendix Table B.1 shows no significant spillover effects: the coefficients on saturation$_d$ are small, mostly negative, and not statistically significant. The coefficients on treated$_d$ $\times$ saturation$_d$ are mostly positive and generally modest in magnitude, indicating treatment effects were not particularly stronger in PMUs with higher treatment saturation.

Table B.1: Spillovers across divisions are minimal

	Contact Survey	
Recognize Round 1 study:	At least 1 reported detail	Number of reported details
	(1)	(2)
	Panel A: Share of divisions assigned to treatment	
Saturation	0.010	-0.051
	(0.046)	(0.116)
Treatment * Saturation	0.061	0.127
	(0.078)	(0.154)
Treatment	-0.035	-0.093
	(0.058)	(0.122)
Observations	578	578
Sample mean	0.04	0.06
	Panel B: Share of divisions treated	
Saturation	-0.008	-0.126
	(0.070)	(0.168)
Treatment * Saturation	-0.011	0.096
	(0.083)	(0.185)
Treatment	0.028	-0.018
	(0.035)	(0.083)
Observations	578	578
Sample mean	0.04	0.06

Notes: This table reports OLS estimates of equation (B.1) using round 1 contact data. Panel A defines saturation as the share of divisions in a PMU assigned to treatment; Panel B, as the share actually treated. Regressions include strata fixed effects by division size and seed survey batch fixed effects. Contacts in divisions with no seed survey are assigned to the control group and a single batch. The dependent variable is: in column 1, a dummy equal to 1 if at least one study feature (e.g., author names, affiliations, journal, or notable elements) was correctly reported; in column 2, the number of features correctly reported. Robust standard errors clustered by division are in parentheses.

B.3 Pre-specified tables

In this section, we discuss results based on the empirical specifications pre-registered in our AEA entry. A few deviations from the pre-analysis plan arose, primarily due to unforeseen data issues.

First, we omit seed-level controls in specifications with the organizational rank treatment indicator. As discussed in Section 3 and following Angrist and Pischke (2009), including these controls would introduce a bad control problem, since senior seeds—likely due to selection—systematically differ in many control variables.

Second, we do not report results for engagement outcomes based on article link clicks. As noted in Footnote 12, a technical issue in round 1 caused some follow-up emails to omit individualized links. Because the issue couldn't be fully reconstructed in Qualtrics, we sent corrected links to all potentially affected seeds. To conservatively address this, we exclude article click data from the analysis.

Third, we report a binary indicator for sharing intentions, collapsing the pre-registered 5-point Likert scale into a dummy: 0 for "very unlikely," "somewhat unlikely," or neutral responses; 1 for "somewhat likely" or "very likely." We present regressions using the original 5-point scale in this section.

We estimate two versions of the tables following the pre-analysis plan, available upon request. The first presents treatment effect estimates with all treatment indicators included in a single regression (as pre-specified), omitting seed-level controls to avoid the bad control problem, but including contact-level controls where applicable. The second version includes only the informational treatments (positive information and credibility), with seed- and contact-level controls where appropriate. The main results in Tables 1 to 3 are robust to both specifications. The results for sharing intentions measured on 5-point Likert scale are directionally consistent, though less precise.

C Survey Details

Seed Survey We randomly selected one employee per division and invited them to complete our survey, introduced as a brief study on AI use at the World Bank. Figure C.1 shows an example recruitment email from round 2 in March 2024. Each selected employee was notified that our aim was to survey one individual per division and had three days to respond. Non-respondents were replaced by another randomly selected employee (provided there were additional eligible employees within the division), up to three times per division. We refer to the selected employees as *seeds*. Surveys in Round 1 were administered via Qualtrics, while those in Round 2 were conducted using SurveyCTO.

Contact Survey We contacted seeds' colleagues 2 to 3 weeks after the final round of seed surveys. We refer to these employees as *contacts* and invited them to complete the "Contact Survey." In round 1, when sending invitations to the contact survey, we reached out to all employees, excluding seeds and those who had previously been invited as potential seeds but did not respond. In round 2, we limited the invitations to employees in divisions where a seed responded and completed the seed survey in either round 1 or round 2.

Employees who respond to our invitation to participate in the round 2 contact survey but work in divisions in which a seed only participated in round 1 of our experiment form the sample for our long-term analysis: these contacts may have learned about the evidence we shared with seeds 3 to 4 months prior to completing the contact survey.

Figure C.1: Sample seed recruitment email

Dear XX,

You are invited to participate in a short survey on Artificial Intelligence at the World Bank conducted by DECRG. We are interested in learning about your views and experiences, no matter your current knowledge about Artificial Intelligence. We aim to survey one employee in each World Bank division. You have been selected at random to participate.

The findings from this study will provide valuable information to the World Bank, and the broader scientific community. Participation is entirely voluntary and your responses are strictly confidential. For more information about the study and data privacy, please see here.

The survey is expected to take **9 minutes** to complete. After completing the survey, you will receive a **$10 Amazon gift card** within a few weeks as a token of our appreciation.

You can access the survey by clicking XX

This link is unique to you. Please make sure to click submit at the end of the survey to ensure your responses are recorded to receive your gift card.

Please note that this survey link will expire on Thursday, March 14 at 6pm.

Thank you in advance! Your support in taking the time to complete this survey is much appreciated!

C.1 Contact Recall Coding

This section describes our procedures for manually coding contact recall outcomes. We asked contacts: Are you familiar with any studies discussing the impact of generative AI tools such as ChatGPT on workers' productivity? In round 2, we replaced "productivity" with "creativity." If they responded "Yes" or said they were familiar with such studies but unsure whether they addressed productivity, we followed up with: You indicated familiarity with studies on generative AI's impact on skilled workers' productivity. To the best of your recollection, please share a few details—e.g., the authors or any notable features. We used these responses to code recall of the studies.

We measured recall of four study features: study authors, institutional affiliation, journal, and content. The authors manually coded contact recall while blind to treatment status. To validate the coding, a research assistant independently replicated the process, also blind to treatment. The authors and research assistant agreed on most cases; disagreements occurred in 2.5% of instances. The authors coded recall as one as one if the correct study authors, institution, or journal were mentioned—even if other studies were cited—and zero otherwise. Disagreement arose only for institutional affiliation: in three cases, the research assistant coded recall as zero while the authors coded it as one because the correct institution appeared alongside others. Given this minor interpretive difference, we used the authors' coding. Study content recall was more subjective; we largely relied on the authors' coding, except in one case of clear error. Notably, our binary variable for recalling any study feature—used in analysis—was almost always coded as one by both coders, even when there was disagreement about specific content, because responses mentioned at least one of the other features.