Labor as Capital: AI and the Ownership of Expertise

This draft: August 14, 2026. First Draft: August 12, 2025.

Zoë Cullen
Harvard

Danielle Li
MIT

Shengwu Li
Harvard

Abstract

Employees generate data about their work, records that can be used to train AI models to perform the same tasks. Combining surveys of 5,000 US full-time workers, stratified across occupations, with a field experiment among knowledge workers at a consulting firm, we document a substantial knowledge gap: much of what workers know about performing their jobs well has yet to be transferred to employers in a form usable for AI training. Second, knowledge supply is elastic. Workers report substantial discretion over what they share, and those experimentally assigned stronger career incentives contribute more knowledge toward training an AI model on their expertise: they write more documentation, share more examples of their work, and disclose more sensitive information, such as client relationships. Motivated by this evidence, we develop a formal model of knowledge supply to AI. When firms own the data by default, workers hold back what they know. Individual rights counteract withholding but set workers against each other, since one worker's contribution weakens the next one's negotiating position. Bargaining collectively restores efficient knowledge supply and allows workers to share in the gains from AI.

JEL Codes: C93, D83, J31, J41, J71.
Keywords: AI, innovation, contracts, labor markets, future of work

Cullen: zcullen@hbs.edu. Li: d_li@mit.edu. Li: shengwu_li@fas.harvard.edu.

This project was reviewed by the Institutional Review Board at MIT. Aaron Leong, Nicolas Miller, Rose Zhang, and Yuerong Zhuang provided excellent research assistance.

1 Introduction

Labor expertise traditionally resides with workers because it involves tacit knowledge developed through experience (Polanyi, 1967). Firms pay workers period by period to access skills they cannot directly possess. In modern workplaces, however, workers increasingly generate data that can be used to train AI models to perform their same work. For example, coding assistants learn from developers' code repositories (Cui et al., 2025; Peng et al., 2023), chat assistants from customer service transcripts (Brynjolfsson et al., 2025), medical assistants from clinical notes (Singhal et al., 2023), and robots from assembly workers' movements (Zhu and Hu, 2018). When human skills are transferred to AI models, expertise is transformed from something firms rent from workers into something they can own outright.

While this transformation can increase productivity, it need not benefit workers. Many AI systems have been trained on knowledge that workers supplied without their awareness or consent, and AI has become a central concern in modern labor disputes (Staff, 2024; Writers Guild of America West, 2023; SAG-AFTRA, 2024; Slator, 2025). At the heart of these debates is a concern about disembodiment: a worker whose core expertise has been captured by an AI system risks losing control over the source of their labor market value. The tension is old. As factories mechanized, Frederick Taylor argued that management should surveil and codify workers' informal know-how, "the principal asset or possession of every tradesman" (Taylor, 1911); workers responded by restricting output, concealing efficient methods, and striking (Montgomery, 1979; Aitken, 1960).

The success of AI deployment depends on access to workers' knowledge of how to perform tasks well.[1] Much of this knowledge is undocumented, residing in the experience and contextual judgment of the people who currently hold these roles. The result is a *knowledge gap*: the difference between what workers know about doing their jobs well and what their employers have captured. This gap is conceptually distinct from AI exposure indices, which measure the share of an occupation's tasks AI could in principle perform; the knowledge gap measures who holds the information required to perform them well. Closing it may require workers to supply knowledge to systems that could one day replace them.[2]

This paper makes three contributions. First, we provide the first measurement of the *knowledge gap* as workers perceive it, and show that it is predictive of how differentiated a workers' tasks are in

[1] A variety of popular press accounts cite the importance of tacit or context-specific expertise. See, for instance, Murty and Kumar S (2026) and Hu (2025).

[2] In some cases, technological advances in self-reinforcement learning may preclude the need for human data(Philip Moreira Tomei, 2026).

their firm, as well as gaps in AI adoption and demand in markets for human expertise. Second, we provide both survey and field-experiment evidence that workers' *knowledge supply* is elastic: workers report many options to withhold or share data about their work, and they supply more data when given stronger career incentives. Finally, we develop a formal model of knowledge supply to AI and show that it may be socially beneficial to give workers the right to bargain, either individually or collectively, over use of their workplace data, rather than allowing employers to own this data by default.

Our empirical evidence comes from two settings: a large survey of full-time US workers (hereafter "US-Occupations Survey"), recruited on Prolific through stratified sampling across economic sectors, and a field experiment within a single firm (henceforth, "the partner firm"). We view these settings as complementary because the sample allows us to have broad occupational coverage and to ask potentially sensitive questions about workers' fears and preferences, while the partner firm sample allows us to observe incentivized in-field worker behavior.

Our partner firm is a large U.S. brokerage and consulting firm that is in the process of building "AI coworkers," digital assistants trained on the expertise of their individual employees. The firm offered nearly all its employees the opportunity to apply to a program focused on building AI Coworkers. Among those who expressed an initial interest, the firm conducted a baseline survey about their job expertise, and then randomized whether employees were offered career incentives. While all workers were reassured that AI Coworkers were not meant to replace them, the treatment group was assigned a stronger set of career incentives internally and externally: 1) the performance of their AI training work could be considered in career advancement and 2) that they would be able to keep and use a copy of their AI Coworker, scrubbed of firm-IP, even if they leave the firm. We observe the results of this baseline survey, whether workers opted to apply to the Coworker training program, and the amount of job knowledge that they actually shared as part of their application process.

We measure the knowledge gap in both the survey and partner firm setting. In both samples, our primary measure asks respondents what share of their workplace knowledge is not captured by their employer's documentation and/or data records. We interpret workers' responses as capturing their *perceptions* of the knowledge gap. Perceptions, whether accurate or not, are important in shaping workers' knowledge supply decisions. Workers report a substantial gap between what they know and what their employers know. On average, US-Occupations Survey respondents report that roughly 40 percent of their job-relevant knowledge is not captured in their employer's documentation

or records. Much of the missing knowledge is tacit and situation-specific: when asked what their employer's documentation fails to capture, workers most often cite knowledge of unusual situations, of people and internal processes, and of general judgment and career skills. These results replicate almost exactly within the partner firm survey (Figure 1). Employees report a mean documentation gap of 43 percent, with most of the undocumented knowledge driven by the same contextual or tacit sources as in the US-Occupations Survey.[3]

Using the survey data (which has wider coverage), we show that our measured knowledge gap correlates with economically important factors in AI deployment at the SOC occupational level. First, occupations with larger average knowledge gaps show larger shortfalls in observed AI usage relative to what frontier models can theoretically perform. This is consistent with practitioner accounts that attribute low AI adoption to a lack of "context".[4] Second, the knowledge gap predicts demand in the emerging market for human-generated training data, where platforms that hire experts recruit more heavily, and spend more, in high-gap occupations. This suggests that firms find it economically worthwhile to fill in knowledge gaps. Third, we show that the knowledge gap relates to workers' bargaining power. Occupations where workers have high individual hold-up power due to differentiated tasks, as measured by Bloesch et al. (2022), are also ones where workers report larger knowledge gaps in our data. In addition to being a technical barrier to AI adoption, this suggests that the knowledge gap is a source of workers' economic rents, and something they may want to preserve.

Next, we provide evidence related to workers' agency over the amount of knowledge they share with their employers. Over 90% of workers in the survey report that they can take concrete actions to either withhold data from their employers or to supply more of it, for instance by communicating off the record or by producing more detailed documentation. They further report that their knowledge supply decisions would meaningfully change the quality of employer records, as measured by how effectively those records could be used to train a replacement.

Our main evidence that knowledge supply is elastic comes from a field experiment with the partner firm. We show that workers who face stronger career incentives are 10% more likely to apply to be a part of the Coworker training program. More importantly, conditional on applying,

[3]Panels C and D of Figure 1 show that the results remain closely aligned when the US-Occupations Survey sample is restricted to occupations similar to the partner firm's workforce.

[4]For example, a recent Boston Consulting Group survey reports that 60% of companies report minimal revenue and cost gains from their AI investments (Boston Consulting Group, 2025), and a global IBM survey of 2,000 CEOs finds that only 25% of AI initiatives delivered expected returns (IBM Institute for Business Value, 2025). On the role of "context" specifically, see Sankar (2025). We choose to use the term "knowledge gap" (rather than "context gap") to include general occupational tacit knowledge that workers may possess, which may be independent of context.

employees are asked to fill in a longer knowledge intake form, with detailed (but optional) questions that they were told could be used to train an AI Coworker. This form asked employees to provide detailed examples and information about their work, with modules focused specifically on experiential and hard-to-codify knowledge: non-standard solutions and workflows they have developed, how work actually gets done inside their organization (how decisions are made and approvals obtained in practice, and who to involve), and, for those in client-facing roles, how they manage their client relationships.

We find that treated workers supply more knowledge: conditional on applying, they answer roughly 25% more questions, write roughly 75% more free text describing their work activities, and share more than twice as many links to work examples and process documentation. These responses are particularly large in knowledge domains that are especially person-specific: information about clients, non-routine workflows, and how things get done within the organization. Consistent with this, we also show the strongest treatment effects are for workers who report having more or more unique knowledge, relative to the firm's documentation. These results suggest that workers have substantial agency in adjusting their knowledge supply when they feel it is more likely to benefit their careers.

Individual agency, however, does not imply individual bargaining power. A worker's ability to profit from supplying their knowledge depends on whether their employer can acquire the same knowledge through other means. Returning to the survey, we ask workers to evaluate the two main alternatives available to firms: expanded surveillance, and seeking data contributions from colleagues doing similar work. Workers report that both strategies can be viable. The average worker believes that fuller workplace monitoring can substantially narrow the knowledge gap and that their colleagues' data contributions are nearly as valuable as their own. This suggests that workers' ability to profit from their data will depend substantially on policies around workplace surveillance and the possibility for collective bargaining.

Finally, in the survey, we ask workers about their preferences over three different data governance regimes: restrictions on the use of work data for AI development, the right to bargain individually over the use of one's work data, and the right to bargain collectively over the use of work data at a firm-occupation level. We find meaningful support for all policies, with the greatest support for individual data rights.

With these empirical findings as motivation, we develop a formal model of knowledge supply to AI, and examine the welfare and distributional implications of three governance regimes: the legal

status quo, in which firms own workplace data; individual data ownership; and collective bargaining over workplace data. One firm and multiple workers interact over two periods. In each period the firm hires workers, who decide how much knowledge to contribute; workers differ in their potential contributions and in their costs of withholding. Because workers are surveilled, the act of working generates records of their expertise: workers supply knowledge by default and must take costly actions to *withhold* it (we later allow workers to incur costs to share additional knowledge).

Knowledge supplied in period 1 becomes training data for an AI system used in period 2. The AI improves the firm's outside option by replicating workers' expertise: even an unfilled position produces output. AI in this model is therefore *expropriative*, whether interpreted as literal automation or as a low-skill hire augmented by the system. We assume workers' contributions are substitutes in training the AI, so that once the firm holds substantial knowledge from some workers, the marginal contribution of others' knowledge is smaller.

Wages and employment are set by Nash-in-Nash bargaining in each period. Two frictions matter. Knowledge contributions are not contractible, reflecting the tacit nature of expertise, so firms cannot simply pay for higher contributions. And period-1 contracts cannot commit parties to period-2 outcomes, so workers face career concerns: contributions today shape bargaining positions tomorrow.

In the baseline model without AI, workers have no reason to withhold knowledge. An equilibrium therefore exists in which all workers are employed in both periods, each contributes their full knowledge, and each receives a share of output proportional to their Nash bargaining weight.

Our first results concern the legal status quo, in which the firm owns worker data. If workers are unaware that their data can train AI models, as many are today, they naively supply full knowledge in period 1. This raises period-2 output but also the firm's outside option, weakening workers' bargaining positions and lowering their wages: the gains from AI accrue primarily to the firm, and worker surplus falls relative to the no-AI benchmark.

Workers, however, are unlikely to remain naive. Once they anticipate how their knowledge may be used, and as long as withholding costs are not prohibitive, there is an equilibrium in which workers withhold knowledge. Relative to naivete, withholding raises worker surplus, but it is socially destructive, reducing total and firm surplus. Withholding is also harder to sustain when workers are more substitutable: when colleagues can provide comparable data, withholding does less to protect a worker's job, so workers withhold less and their wages fall as expropriation increases. This comparative static is what the firm data show in cross-section: supply rises with perceived

substitutability, and the workers who hold back, restrict access, and fear displacement are those whose knowledge colleagues cannot replace.

We use this model to shed light on several alternative policies: banning the use of work data for AI training, granting workers individual ownership over their work data, and establishing collective ownership of work data among workers.

Banning the use of work data for AI training, in our setting, restores the baseline no-AI equilibrium, improving workers' job security, but forgoing potential productivity gains from AI.

We model individual ownership as the right to bargain over whether, and at what price, one's data is used at time-2; in the event of disagreement, the firm cannot use it. Ownership lets workers profit from their work data and removes the motive to withhold. It does not, however, guarantee full employment. Each worker's data improves the firm's bargaining position vis-a-vis every other worker, lowering others' future wages; because hiring a worker adds their data, hiring can raise colleagues' wage bill by more than it raises output, and we give an example in which the firm responds by freezing hiring—leaving a worker unemployed and total surplus strictly below the no-AI benchmark. Under a condition on the AI technology that rules out such threshold-crossing effects, individual ownership supports an equilibrium with full employment and full contributions, and maximizes social welfare. Even so, individual ownership need not translate into greater pay: the externality just described is stronger the more substitutable the data, and in the extreme case when workers are symmetric and their data are perfect substitutes, workers receive *none* of the output gains from AI, no matter their Nash bargaining weight.

The last policy we consider is one in which workers pool their work data in a collectively governed data trust: the trust bargains with the firm over the use of the pooled dataset (in the event of disagreement, the firm cannot use any worker's data), while employment and wages remain individually negotiated. Like individual ownership, collective data ownership restores workers' knowledge contributions to the social first best. However, collective data ownership also eliminates the competition externality present under individual ownership, because one worker's contribution no longer improves the firm's bargaining position against other workers. As a result, workers capture a larger share of the surplus generated by AI. Moreover, the firm is always weakly better off under the data trust than under the no-AI benchmark, and each worker's compensation for the use of their data—their time-2 compensation—weakly exceeds its no-AI level for every trust bargaining weight; neither guarantee requires workers to bargain collectively over wages: collective control of the data alone suffices. The trust's protection has a limit, however. Because time-1 wages remain individually ne-

gotiated, a worker whose colleagues' data can already replace them has a weak bargaining position at the hiring stage, and the firm can extract part of their future data rents there, so workers' *total* payoffs are not guaranteed to exceed the no-AI benchmark. While our simple model abstracts from frictions within the trust, it nonetheless suggests that collective action could be a useful policy tool to help workers share in the gains from AI.

Two further results sharpen the comparison with individual ownership. First, whenever the trust's bargaining weight over data is at least the workers' individual bargaining weight, total time-2 payments to workers are weakly higher under the trust: bilateral bargaining pays each worker for the *marginal* contribution of their data, and when data are substitutes the sum of marginal contributions falls short of the value of the bundle that the trust sells. Second, the compensation guarantee does not depend on the trust bargaining well: each worker's time-2 compensation weakly exceeds its no-AI level for *every* trust bargaining weight. And in the hiring-freeze example described above, the data trust restores full employment and attains the first best.

Finally, motivated by our empirical evidence, we consider an extension in which workers can also contribute *more* knowledge at a cost. The firm's program operates on exactly this margin: contributing expertise requires active, costly effort, and under firm ownership the extension predicts the pattern we observe, incomplete contribution that is thinnest among workers with the least substitutable knowledge. Under individual ownership, by contrast (maintaining the condition that rules out hiring freezes), workers exert effort to codify their expertise so that they can be compensated for the resulting AI improvements. The welfare effects are ambiguous, since workers may exert more than the socially efficient effort; under the data trust, total surplus is always at least as high as with no AI.

Overall, despite the popularity of individual ownership among our survey respondents, our model suggests three advantages of collective data ownership. First, the data trust protects workers' compensation for their data: each worker's time-2 compensation weakly exceeds its no-AI level however weak the trust's bargaining position, whereas individual ownership can leave every worker strictly worse off than no AI. Second, the data trust sustains full employment for any AI technology consistent with our assumptions, whereas individual ownership can lead the firm to freeze hiring, sacrificing employment and total output. Third, when workers can contribute additional knowledge at a cost, the data trust always improves *total* surplus compared to no AI, whereas individual ownership sometimes does not.

These findings bear on the governance of AI-augmented production. Policy debates focus on consumer data and privacy; workplace data raise distinct questions about the ownership of surveilled human capital. Worker awareness without data rights may reduce productivity as workers withhold knowledge, and individual rights alone may not raise worker welfare, because workers do not internalize the effect of their data sales on colleagues. Collective institutions better align private incentives with social welfare.

2 Background

2.1 Related Literature

A growing literature measures AI's impact on productivity and labor, once deployed (Brynjolfsson et al., 2025; Peng et al., 2023; Cui et al., 2025), but there has been less work on the question of how AI models build expertise. Our paper is among the first to highlight the importance of labor in supplying the knowledge necessary to build AI capital. A small set of papers have begun to examine the implications of this premise: Brynjolfsson and Hitzig (2025) argue that AI shifts decision rights toward the center of organizations, both by codifying tacit local knowledge and by relaxing the limits on central information processing; Chequer et al. (2026) build a macro-labor model in which workers generate the in-house data firms need to adopt and improve AI, and show that labor remains valuable even under aggressive automation. This paper is distinct in that it focuses on worker agency over the work data they supply, and its implications for welfare, wages, and data ownership policies.

We introduce two new objects: the knowledge gap, the stock of job knowledge that workers hold and their employers do not, and knowledge supply, workers' choice of how much of that knowledge to reveal. In doing so, we treat workplace knowledge as distinct from human capital, and knowledge supply as distinct from labor supply. A long literature recognizes tacit and firm-specific knowledge as inputs into current productivity: workers know more than they can tell (Polanyi, 1967), knowledge specific to a firm shapes wages, turnover, and rent sharing (Becker, 1964; Lazear, 2009), and evidence from tenure profiles, displacement, and worker deaths indicates that such knowledge is quantitatively substantial (Shaw and Lazear, 2008; Jacobson et al., 1993; Jäger and Heining, 2022). In that literature the distinction between knowledge and labor carries little weight because knowledge is embodied: it raises output only through the labor of the worker who holds it,

so supplying one means supplying the other. This equivalence breaks down in our setting, where knowledge can be preserved and embedded in AI models.

Our model concerns the strategic supply of knowledge. The friction driving it, that revealing knowledge today worsens the terms a worker faces tomorrow, is shared with the ratchet, hold-up, and career-concerns traditions (Freixas et al., 1985; Gibbons, 1987; Grout, 1984; Holmström, 1999), but we introduce two new elements. First, revealed knowledge accumulates in a durable asset, an AI system in the firm's disagreement point, rather than in a principal's beliefs, which puts relational remedies out of reach (Baker et al., 1994) and makes data rights the relevant instrument. Second, revelation is multilateral: substitutable contributions mean each worker's data weakens every colleague's bargaining position, a contracting externality in the sense of Segal (1999) (see also Stole and Zwiebel, 1996). Our comparison of ownership regimes is then in the spirit of Grossman and Hart (1986), with collective ownership emerging as the arrangement that protects non-contractible contributions while internalizing the externality.

Finally, our analysis connects the economics of data (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2021; Bergemann and Bonatti, 2019) to labor market outcomes. That literature has centered on consumers: individuals sell information about themselves, and externalities from data sales are informational, running through privacy (Acemoglu et al., 2022; Acquisti et al., 2016). In contrast, our data suppliers are workers, and the externality runs through wages, because one worker's data trains a substitute for colleagues' knowledge. Work data also arise jointly with production and, under current law, are extracted by default through monitoring rather than sold by choice. These differences are why individual data rights, a natural remedy in consumer settings, can fail workers here, and they reshape the case for collective institutions. Data collectives have been proposed as a way to pool consumer bargaining power against platforms or to manage privacy externalities (Posner and Weyl, 2018; Arrieta-Ibarra et al., 2018); our results provide a labor-market rationale.

2.2 Worker Data and AI Development

Recent advances in AI, particularly large language models (LLMs), rely on both public and proprietary data. Foundation models are trained on public data (internet text, books, and code) to develop general capabilities, but they often lack the domain-specific knowledge needed for particular organizational tasks.

Bridging this gap requires adapting models with proprietary data (customer support logs, call transcripts, internal documentation), often generated by workers in the course of performing these tasks. These data capture not only formal procedures but, in some cases, knowledge that was previously tacit: skills and heuristics observable in practice but difficult to codify (Polanyi, 1967).

The following examples illustrate how worker-generated data can be used to build AI systems:

Call Recordings and Customer Service AI: Customer service models are trained on call transcripts and audio logs. In many cases, these transcripts are labeled not only with objective performance metrics such as call duration but also with indicators for whether the text was generated by a top-performing agent. Such labels allow the model to mimic the skills of specific individuals (Brynjolfsson et al., 2025).

Movement Video and Robotic AI: Video of humans performing physical tasks is used to train robotic systems. For example, NVIDIA's foundation model for humanoid robots is trained in part on video of human demonstrations, and firms now pay workers to record themselves performing tasks such as folding laundry, with wearable cameras capturing their hand movements as training data (NVIDIA, 2025; Christopher, 2025).

Clinical Notes and Medical AI: Clinical notes written by providers are used to fine-tune medical AI systems. For example, NYUTron, trained on years of clinical documentation from a single health system, is used to predict patient outcomes, generate clinical reports, and support discharge decisions (Jiang et al., 2023).

Task and workplace specific data can be captured in a variety of ways. Most notably, workplace monitoring is now pervasive and produces durable records of how labor is performed: a 2024 OECD survey finds that about 90% of U.S. firms monitor workers in some way, 75% collect data through surveillance tools, and most do not allow workers to opt out (Milanez et al., 2025; U.S. Government Accountability Office, 2024; Hertel-Fernandez, 2024). Policy attention has focused on monitoring's direct effects (privacy, safety, anxiety) with little notice that the same records can be repurposed to train AI systems that may ultimately perform the monitored tasks.[5]

In addition, workers increasingly have opportunities to be paid to supply their work knowledge. Platforms like Mercor explicitly recruit professionals such as doctors, lawyers, and financial analysts, paying over $250 per hour at times for their specialized domain knowledge (Wall Street Journal,

[5] U.S. Government Accountability Office (2024); Milanez et al. (2025); Hertel-Fernandez (2024) document concerns related to autonomy, privacy, and workplace safety, but do not mention the use of surveillance data for AI training.

2026). These markets are no longer niche: between January and July 2026, data posted on the job board of a leading talent-sourcing platform, which AI companies use to source expertise to train their models, show postings across essentially every major occupation group, including law, medicine, finance, consulting, engineering, education, and the trades, with posted rates of $50–$250 per hour and thousands of workers hired per month (see Section 3).

Finally, debate continues over whether AI will achieve human-level general intelligence (Grace et al., 2022; Allyn-Feuer and Sanders, 2023). While definitions of artificial general intelligence (AGI) emphasize broad reasoning and problem-solving capabilities (Bubeck et al., 2023), even highly capable models would still require context-specific training and examples to excel at actual workplace tasks (Mitchell, 2021). Although advanced systems may learn such knowledge autonomously, it is likely more efficient to provide models with examples from human experts who already possess this information. As a result, human-generated data and experience are likely to remain important inputs to AI systems (Ramani and Wang, 2023).

2.3 Institutional and Legal Context

The prevailing legal framework in the United States gives employers broad ownership over workplace data. Work product created within the scope of employment generally belongs to the firm under the work-for-hire doctrine, and monitoring records collected on employer systems are typically treated as firm property as well (Kim and Leavitt, 2026; U.S. Congress, 1976).

New AI capabilities are putting pressure on this framework and prompting legal and labor responses. In its 2023 contract negotiations, the Screen Actors Guild secured provisions barring studios from using recordings of an actor's likeness to create digital replicas without consent and compensation (SAG-AFTRA, 2023). More broadly, researchers and advocates have proposed governance frameworks, such as collective data rights and worker data trusts, intended to give workers a share in the value their data creates (Viljoen et al., 2026; Kim and Leavitt, 2026; Diamantis, 2023).

These debates point to a gap in the evidence. AI systems rely on worker-generated data, yet little is known about how much such knowledge workers hold, how willing they are to supply it, or what policies they would prefer to govern it.

3 Empirical Evidence on the Knowledge Gap

We measure the knowledge gap by conducting a large, stratified survey of full-time U.S.-based workers recruited through the survey platform Prolific. The survey collects information on participants' work, with a focus on job-specific expertise they possess, as well as their ability to influence how much of this knowledge can be recorded or documented by their employer.

3.1 Online Survey Sample

We recruited 4,043 workers who were currently employed full time through Prolific. The survey questions were designed around each respondent's primary job. Recruitment was stratified across economic sectors on Prolific: we set sector-level quotas so that the sample spans the full occupational distribution of U.S. full-time work.[6] Within strata, we construct post-stratification weights so that the weighted sample matches the Census distribution of full-time workers on demographics (sex, age, and education) within major (SOC 2-digit) occupation groups. Unless noted otherwise, figures and tables report weighted statistics; unweighted results are similar.

Appendix Figure A1 illustrates the breakdown of respondent occupations, with the most common being management (16.2%), business and financial operations (12.5%), office and administrative (12.3%), education and library (8.0%), and computer and mathematical (7.7%). Common job titles include manager, teacher, and software engineer. The median subject is 37 years old; 40% of subjects are male and 69% are White. We include a comparison between our sample and a cross-section of the U.S. full-time workforce in Appendix Table A1.

3.2 Partner Firm Sample

Our second setting is the partner firm, a large U.S. brokerage and consulting firm whose employees advise client organizations on employee benefits, insurance, retirement and wealth planning, payroll, and human resources. The firm is building "AI Coworkers," digital assistants trained on the expertise of individual employees, and in mid-2026 it invited nearly all of its roughly 5,300 employees to apply to a program dedicated to building them. Employees who expressed initial interest completed a baseline survey about their job expertise; the 821 employees (about 15 percent of those eligible) who completed the survey through the program-application decision form our main analysis sample.

[6] Our resulting respondent population is more representative of U.S. workers, although it still over-represents workers in computer-based occupations because of Prolific's respondent population.

The sample spans the firm's functional teams, from benefits consultants, account managers, and client service to retirement, wealth, analyst, and leadership roles (Appendix Figure A1, Panel B).

The baseline survey mirrors the survey instrument. It asks the same documentation question (what share of the respondent's work knowledge is captured by the firm's training materials, knowledge databases, role documentation, and examples of past output), the same unique-knowledge question (what share is captured by neither the firm's documentation nor colleagues' knowledge), and the same multi-select question about the types of knowledge that documentation misses. It also records each employee's familiarity with the AI Coworker initiative and their pre-existing intent to create one, which we use as controls in the treatment-effect analysis. Because respondents are employees rather than anonymous panelists, we can link their responses to administrative records: the employee roster (title, practice, division, tenure, and reporting relationships) and individual usage metrics from the firm's internal AI platform.

After the baseline questions, the survey randomized the incentives under which employees decided whether to apply. Employees in the treatment arm were shown a set of internal *and* external career incentives for creating an AI Coworker: their training work would be visible to their manager and could be considered in career advancement; the Coworker they train could be recognized through broader internal deployment and a company award; and they would keep a copy of their Coworker, scrubbed of firm intellectual property, even if they left the firm. The control arm received the same invitation without these inducements. Employees in both arms were assured that AI Coworkers were meant to augment rather than replace them. The incentive treatment was randomized across clusters of employees defined by sub-trees of the firm's reporting hierarchy, so colleagues who work together saw the same version; all treatment-effect estimates cluster standard errors on this randomization unit. Treatment and control arms are balanced on roster characteristics, baseline intent, AI familiarity, and pre-treatment AI-platform use.

We measure knowledge supply at two margins. The extensive margin is the application decision itself. Applicants were then directed to a knowledge-capture form with detailed but optional free-text questions about their work: how their approach to specific tasks deviates from standard practice and is hardest to explain, how decisions and approvals work in their part of the organization, what they know about specific clients, and what they know about their market and industry. The form also asked for links to examples of their work output and documentation of their work processes on the firm's internal systems. Respondents knew that their answers could be used to train an AI Coworker with their expertise. From this form we measure the share of questions answered, the

volume of free text written, the number of links shared, and the number of tasks described, and we combine the four into an equal-weighted standardized supply index. Employees who did not complete the form are coded as supplying zero, so the index reflects both the decision to participate and the effort supplied conditional on participating.

3.3 Measuring the Knowledge Gap

Our knowledge gap measure captures knowledge that workers possess but that their employers currently do not. We ask respondents to consider the full range of information their employer holds about their role, spanning both "official" and "unofficial" sources. Official sources are categories of data already codified and recognized as job knowledge: policy manuals, knowledge bases, training materials, and AI assistants. Unofficial sources are data the firm may be gathering but has not (yet) formally codified as job knowledge: computer activity logs, communication records, performance metrics, and work outputs. Such records are pervasive. As Appendix Figure A3 shows, a majority of respondents report that their employer monitors their performance (75%), communications (70%), outputs (62%), and time-use (55%). Substantial minorities report that they are subject to video (44%), location (44%), computer (40%), or audio recording (27%). Taking this full record as given, we ask respondents what share of their work knowledge it captures. Because we credit the employer with this broad set of holdings, our definition of the knowledge gap (what workers *additionally* possess) is intended to be conservative. Finally, we also ask respondents to think about the specific aspects of their knowledge that are not well captured by their employer's documentation.

Our first measure defines the knowledge gap in terms of *documentation*: what share of a worker's job-relevant knowledge is not captured by their employer's documentation or records? On average, workers in the US-Occupations Survey report that around 40% of their workplace knowledge is not held by their employers, and partner-firm employees report a nearly identical mean gap of 43%. Panel A of Figure 1 shows the distribution of this gap among US-Occupations Survey respondents, and Panel B reports the aspects of work knowledge they most often say are not well captured by their employer's documentation. The most common are knowledge of unusual situations (64%), people and internal processes (54%), and general judgment and career skills (49%), suggesting that workers hold a lot of tacit and situationally-specific knowledge. Panels C and D repeat both exercises for the partner firm, comparing its employees with US-Occupations Survey respondents in occupations matching the firm's workforce (Appendix Figure A1 reports sample composition): both the distribution of the gap and the ranking of missing knowledge types line up closely across the

two samples. Split across SOC-2 occupation codes (Appendix Figure A4, Panel A), respondents in the arts and education fields report the largest average gaps (near 50%), while respondents in the transportation field report the smallest (around 30%). Within-occupation variation is greater still, with an interquartile range spanning roughly 40 points in the typical occupation; Panel B of that figure shows a similar pattern across functional teams within the partner firm.

A second measure converts the knowledge gap from shares of documentation into shares of job *performance*: we ask respondents to estimate the performance of a hypothetical replacement with similar experience and general skill, trained only on the firm's documentation and records. For sensitivity reasons, this question could be asked only in the survey, not of the partner firm's employees. Workers' responses indicate that they believe the documentation gaps shown in Appendix Figure A4 have meaningful consequences for performance: on average, workers report that such a replacement would perform their job only about half as well as they do. The occupational ranking is similar to Panel A of Appendix Figure A4, with arts sitting at the top of both rankings, and transportation at the bottom. Appendix Figure A5 shows the person-level distribution of this measure and that the two measures of the knowledge gap are highly correlated at the individual level. Because it is available in both samples, we use the documentation elicitation as our primary quantitative measure of the knowledge gap; the performance elicitation serves as a complementary measure of the gap's productivity consequences.

A knowledge gap indicates what employers do not possess, but not who does. Knowledge missing from the firm's records may be shared widely among colleagues, or may reside with a single worker. We therefore ask workers what share of the knowledge gap colleagues could not fill. The average worker reports that roughly half of their uncaptured knowledge is theirs alone, though the distribution is wide, with an interquartile range from 21% to 75%. This distinction is the margin that matters for bargaining: knowledge held in common gives no individual worker leverage, while uniquely held knowledge does. Figure 2 plots the distribution of this uniquely held knowledge, the share of work knowledge captured by neither employer documentation nor colleagues, in both surveys: partner-firm employees report a mean unique gap of 32 percent, compared with 23 percent among respondents.

3.3.1 Interpretation

The average worker reports that roughly 40 percent of their job-relevant knowledge is not captured in their employer's records, and that a similarly experienced replacement limited to the employer's

documentation would perform at roughly half their own level. These gaps are large and may in part reflect workers' potentially inflated perceptions of their own knowledge and abilities.[7] The close agreement between the anonymous responses and those of partner-firm employees, who answered about a specific, named employer inside their own workplace, suggests the measure is not an artifact of either survey environment.

The reported gap is also consistent with what direct measurement shows when workers are actually replaced. The closest available benchmarks come from tenure-output profiles: among new hires at an autoglass company, output rises by roughly 50 percent over the first year, so recent hires begin at about two thirds of an experienced worker's level (Shaw and Lazear, 2008); Brynjolfsson et al. (2025) find a similar figure for call center workers. Other estimates are measured in different units but point to gaps of a comparable order of magnitude. Long-tenured workers who lose their jobs suffer persistent earnings losses averaging 25 percent per year (Jacobson et al., 1993) and unexpected worker deaths generate substantial replacement costs that rise with skill specialization (Jäger and Heining, 2022). These benchmarks differ from our survey scenario in two offsetting ways. Actual replacement hires are typically less experienced than the incumbents they replace, while our question holds experience and skill fixed; on this dimension, observed gaps should overstate the knowledge gap we ask about. However, actual replacements also learn from coworkers and managers rather than from documentation alone; on this dimension, observed gaps should understate it.

Regardless of levels, workers' perceived knowledge gap varies across workers and firms in ways that theories of job-specific knowledge would predict. Appendix Figure A6 plots the knowledge gap against worker and firm characteristics in the sample, comparing workers within the same major occupation group. Gaps rise with education and pay (Panels A and C). Years in the current role show little gradient in the sample (Panel B), but career-long experience in one's field, which we observe for partner-firm employees, is strongly associated with the gap: employees with less than a year in their field report gaps around 30 percent, against nearly 50 percent for those with more than twenty years (Appendix Figure A7, Panel B). This pattern is consistent with tacit knowledge accumulating over a career rather than within a single position. In addition, Panel D shows that the knowledge gap varies reasonably with firm choices: controlling for education and income, workers whose employers keep richer records claim less knowledge beyond those records. While these correlations do not

[7] For example, (Hoffman and Burks, 2020) show that workers over-predict their own productivity in field data. The magnitude of overprediction, 10 to 25 percent, if applied to our data would still generate a large knowledge gap.

rule out some inflation in our measure of the knowledge gap, they suggest that there is information content in these worker reports.

3.3.2 Knowledge Gap and AI Adoption

The knowledge gap is conceptually and empirically distinct from AI exposure. Exposure measures capture the share of an occupation's tasks that AI could technically perform (Eloundou et al., 2023); the knowledge gap measures whether the employer holds the information an AI system would need to perform them. Figure 3 plots occupation-level knowledge gaps against theoretical AI exposure at the SOC-3 level, with dashed lines at sample medians. The two are only moderately related, and occupations populate all four quadrants. Occupations like record clerks and services sales representatives combine high exposure with a low knowledge gap: their work is both automatable and well documented, making them candidates for effective AI adoption. Legal support, secretaries, and media/communication (writer) roles are equally exposed but report high knowledge gaps: AI capability exists on paper, but deployment depends on capturing knowledge that firms currently do not hold. Motor vehicle operators and construction occupations combine low exposure with a low knowledge gap, suggesting that AI adoption there awaits general technology improvements, such as robotics, rather than worker-supplied data. Food service occupations are low exposure but high knowledge gap: automation may require both technological gains and the collection of more worker data (Yang and Zhou, 2026).

This reasoning implies that where knowledge gaps are large, AI adoption should lag its technical potential. Figure 4 examines this relationship.

Using data from Massenkoff and McCrory (2026), we construct an occupation-level "Missing AI Index": the difference between an occupation's theoretical exposure to AI and its observed use, measured from Anthropic usage logs. Theoretical exposure captures the share of occupational tasks that an LLM could plausibly automate or accelerate; observed exposure reflects only those tasks that actually appear in work-related Claude interactions. Their difference measures the extent to which conceptually feasible AI use has not materialized in practice. Figure 4 shows that occupations with a high measured knowledge gap are also ones with a greater gap between potential and actual AI usage: a one standard deviation increase in the knowledge gap is associated with a 0.37 standard deviation increase in the AI use gap. We observe a similar pattern within the partner firm: employees who report larger knowledge gaps make less use of the firm's AI platform (Appendix Figure A7, Panel A).

If uncodified knowledge is what stands between theoretical AI capability and realized use, we should expect a market to emerge for exactly that knowledge. Figure 5 displays the positive relationship between demand for human expertise to train AI and the self-reported knowledge gap at the minor occupation level (SOC-3). In this figure, demand for human expertise is measured using daily observations from a leading job platform for AI training, including the hourly pay rate and number of people hired for each role.[8] Occupations whose workers report larger replacement gaps see more expert hiring for AI codification ($\rho = 0.38$ on log hires; 0.37 using the knowledge gap measure): the same occupations our respondents identify as running on uncaptured knowledge, such as arts and media, and computing, are those where AI developers are paying to acquire it. The marketplace thus provides revealed-preference validation, from the demand side, of what our survey measures from the supply side. Appendix Figure A8 Panel A shows that demand appears first in domains with high levels of documentation, including material recording, other education/library roles like teaching assistants, and sales, and later in domains with more limited documentation including law, secretarial work, and media/communication (which includes writers), consistent with codification proceeding up the tacitness gradient documented above. Panel B and C show that despite a weak correlation with the hourly posted rate, the overall spend in the market for expertise continues to be positively correlated with our measure of the knowledge gap.

Finally, the knowledge gap relates to workers' bargaining power. Figure 6 plots the occupation-level task-differentiation measure of Bloesch et al. (2022), a measure of workers' individual hold-up power based on the dissimilarity between a worker's tasks and those of their coworkers, against our knowledge gap measure. Occupations where workers report larger knowledge gaps are also occupations where workers hold more individual hold-up power ($\rho = 0.44$). Beyond acting as a technical barrier to AI adoption, the knowledge gap is thus also a plausible source of workers' economic rents, and something they may want to preserve.

3.4 Knowledge Supply

3.4.1 Worker Actions

Having provided evidence that the knowledge gap is potentially important for productivity, we next examine workers' scope for narrowing or widening this gap.

[8] Postings are classified to a detailed SOC occupation using large language models and aggregated to the SOC-3 minor group. For details on construction, see (Cavallo and Cullen, 2026).

Workers report near-universal discretion over their knowledge supply: 90% identify at least one action they could take to reduce the data their employer collects about their work, and 96% identify at least one to increase it. Panel A of Figure 7 presents the share of workers reporting each action. The most common channels on both margins are ordinary communication and documentation practices rather than technical ones.

Workers believe these actions have meaningful consequences for their replaceability. Panel B of Figure 7 reports workers' beliefs about the knowledge gap facing a similarly experienced replacement, under alternative knowledge supply regimes. Against a baseline gap of 45%, workers report that individually withholding all the data they could would widen the gap to 54%, while supplying all the data they could would narrow it to 21%. Workers' incentives to share knowledge therefore matter most where the baseline knowledge gap is large.

3.4.2 Field-Experiment Evidence on Knowledge Supply

The field experiment at the partner firm shows this agency in action. When the firm attached stronger career incentives to sharing expertise, workers shared more, on every margin we can measure. Workers randomized into stronger career incentives are more likely to apply to the AI Coworker training program: applications rise by 6.2 percentage points on a control-arm base of 64% (top row of Figure 8; Table 1), a roughly 10 percent increase in participation from incentives that were modest and non-pecuniary: visibility with one's manager, internal recognition, and a portable copy of the resulting AI Coworker.

The richer margin is how much knowledge workers turned over once they applied. Conditional on applying, treated workers also supply more knowledge: they answer more of the questions they are shown (about a quarter more, relative to a control mean of 34 percent), write more free text describing their tasks and working relationships (0.55 log points), and share more than twice as many links to work examples and process documentation (Table 2, Panel A). Combining the four outcomes into a standardized index, the career-incentive treatment raises knowledge supply by 0.25 standard deviations ($p < .01$) among appliers. The estimate is nearly identical in the full randomized sample with non-appliers coded as supplying nothing (Panel B), so the effect reflects genuinely greater supply rather than selection into applying.

Nor is the additional knowledge confined to easily codified material. Table 3 breaks the effect out by the type of knowledge supplied: treated workers answer 7 to 8 percentage points more of each free-text question family, on control means around 20 percent, with similar responses for

non-standard practices, internal organization, client relationships, and market knowledge. The tacit, context-heavy categories that both surveys identify as least documented respond at least as strongly as the rest, and Figure 8 summarizes the treatment effects across outcomes and content areas as a percent of the control-arm mean. The incentive effect also rises with the worker's initial knowledge gap and with the uniqueness of the worker's knowledge (Figure 9): the treatment effect on the supply index grows by about 0.007 standard deviations per point of the initial knowledge gap ($p = .016$), and the quartile estimates show the response concentrated among workers in the upper half of each distribution (Panels B and D). The unique-knowledge pattern is especially telling. In the control arm, supply *declines* with uniqueness: holding the size of the total knowledge gap fixed, workers whose knowledge colleagues cannot replicate supply about 0.0045 standard deviations less per point of uniqueness, consistent with strategic withholding by workers who understand that unique knowledge is a source of leverage. The career-incentive treatment offsets this gradient (interaction 0.0049, $p = .11$), so that treated workers supply at similar rates regardless of how unique their knowledge is (Panel C). The largest impacts of the treatment thus fall on the workers with the most reason to believe their knowledge is valuable: those with the largest gaps and the most unique knowledge. Supply responds most precisely where the previous sections suggest supply matters most.

Taken together, the experiment documents the fact our framework requires: knowledge supply is elastic. A modest shift in career incentives moved both participation and the quantity and content of the knowledge that employees turned over, knowing it could train an AI system on their work.

3.5 Limits to worker agency

Firms facing worker agency may respond by creating incentives for a given worker to share their expertise, as our partner firm did, but they may instead choose to limit individual worker agency by acquiring the expertise without that worker's cooperation. We focus on two versions of the second approach: greater surveillance and data from other workers. Neither can be randomized within a single firm, so we return to the US-Occupations Survey sample, whose counterfactual questions ask workers to assess exactly these scenarios for their own jobs.

Panel A of Figure 10 reports workers' perceptions of the maximum knowledge their employers could acquire through surveillance. Specifically, we ask them to "imagine that your employer fully monitors your work by collecting data on everything you write, say, or do at work (but not your inner thoughts or decision-making process)." Relative to what employers currently document, workers

believe there is substantial scope for firms to capture knowledge through surveillance. The average worker reports that full monitoring would narrow the knowledge gap facing a replacement to around 20%, the same as what workers believe their own affirmative data contributions could achieve (Figure 7, Panel B).

Firms vary in the degree to which they may be able to reduce the knowledge gap through additional surveillance. Panel A of Appendix Figure A9 shows that the returns to full monitoring are largest where current monitoring is lightest: within the same occupation and industry, workers at firms that do not monitor report nearly three times higher returns to full surveillance than workers at firms that already capture much of their data (34.3 versus 13.3 percentage points). The knowledge gap a firm faces is thus partly determined by its own choices, not just the nature of a worker or their job.

In practice, it may be difficult to fully surveil workers, or to incorporate all types of surveillance records into AI training. An alternative to relying on data contributions from a given worker is to seek contributions from other workers performing similar work. Panel B of Figure 10 presents workers' perceptions of how the knowledge gap for their own job would shrink under two scenarios: if their colleagues did everything they could to document their work (but the respondent did not), or if the respondent did everything they could to document their work (but their colleagues did not). On average, workers believe that colleagues' data are a good substitute for their own data, and that both can reduce the knowledge gap to around 20-25%, similar to the knowledge gap under full surveillance.

Panel B of Appendix Figure A9 shows that the value of colleagues' data can vary across workplaces. Workers with no colleagues in a similar role rate colleague data roughly 13 percentage points less valuable than their own. Among workers with at least one similar-role colleague, however, colleague data is worth nearly as much as their own, and its relative value is flat in the number of colleagues: access to even a small number of close colleagues erodes the uniqueness of a worker's knowledge. We find the same pattern among partner-firm employees, whose reported unique knowledge declines with the number of colleagues doing similar work.

Together, these results suggest that although workers report considerable agency over their own knowledge supply, firms may also have considerable scope to acquire worker knowledge without any individual worker's cooperation. As surveillance expands, passive data flows displace voluntary contributions as the primary channel of knowledge capture, and workers' ability to withhold, rather than their willingness to share, becomes the more important margin of agency. And when colleagues

can supply substitute data, no worker's supply decision is independent: one worker's contribution erodes the value of another's withholding, so individual data decisions carry externalities that only coordination can internalize. Both forces raise the stakes of collective action. We return to these issues in our model.

3.6 Policy Preferences

We consider worker demands for regulation of labor data, which fall into three broad categories:

Banning the use of AI in the workplace: The most restrictive demand prohibits specific AI or automation uses altogether. The International Longshoremen's Association's master contract, secured after a three-day coast-wide strike in October 2024 and ratified in February 2025, bars AI and automated systems from certain longshore functions and requires employer negotiation and worker consent before new technology is deployed (International Longshoremen's Association, 2025; USMX and ILA, 2025). In creative work, the Writers Guild of America's 2023 Minimum Basic Agreement stops short of an outright ban but reserves the Guild's right to assert that using writers' material to train AI is already prohibited under the agreement or other law (Writers Guild of America West, 2023).

Mandating individual consent: This demand accepts AI use but conditions it on the informed consent of the worker whose data is used. SAG-AFTRA's 2025 Interactive Media Agreement, ratified by 95% of voting members after an eleven-month strike, requires separate, "clear and conspicuous," reasonably specific written consent before a digital replica of a performer's voice or likeness may be created or used, and permits performers to suspend that consent during a strike (SAG-AFTRA, 2025).

Mandating collective consent: This demand moves decisions about data use from the individual to a collective structure governing pooled data under shared rules. Driver's Seat, a driver-owned cooperative, enabled ride-hail and delivery workers to pool the trip, time, and earnings data they generated across gig platforms, with members sharing in profits and electing the board (Orchi, 2023). The cooperative sold this pooled data (for instance, to local and state governments for transportation planning) so that value flowed back to the drivers who

produced it (New Jersey Office of Innovation, 2026). Workers with little individual leverage over their data could thus collectively govern its use and bargain over its sale.[9]

We ask a random subset of our sample of full-time US workers on their preferences over these policies related to their labor data. We present experimental vignettes to our treated subjects, describing three alternatives. The first policy bans the use of passively-collected work data in AI development[10]; the second allows for individual control over one's own work data, including its sale[11]; and the third allows for collective worker control and sale of work-data.[12]

To assess workers' preferences, we follow the methodology of Buckman et al. (2025).[13] Panel A of Figure 11 displays the share of workers who indicate that they like each policy. While there is net positive support for all three options, individual ownership of work data emerges as the most popular choice. Approximately 63% of respondents favored a policy that would grant them the right to own and sell their individual work data for AI development. Panel B plots workers' relative support for collective versus individual ownership against the number of colleagues they report working in a similar role, residualized by workers' broad occupation category. Workers with five or fewer similar colleagues favor individual ownership by a wide margin, while workers with more than one hundred lean toward collective ownership. This pattern anticipates the logic our model will later formalize: the returns to collective ownership are greater when colleagues' data can replace one's own.

4 Theory

4.1 Overview

We develop a model of AI expropriation to understand how worker awareness of surveillance affects knowledge sharing and welfare outcomes.

[9] Driver's Seat has since been discontinued; the Workers' Algorithm Observatory at Princeton notes the project has been "sunsetted," with its mission carried forward through the FairFare tool (Workers' Algorithm Observatory, nd).

[10] "Under this policy: 1. Employer could not use work data to develop AI models, or sell work data to other firms to develop AI models'. 2. Those seeking to develop AI models would have to hire workers to specifically produce AI training data; they could not use data that was produced as the byproduct of every day work tasks."

[11] "Under this policy: 1. You decide if your data can be used for AI training. You could say no, and your employer couldn't include your data in their AI models. 2. Your employer can pay you to use your data, at a price that you both agree on."

[12] "Under this policy: 1. You and other workers in your role jointly decide whether your collective work data can be used for AI training. You could collectively say no, and your employer couldn't include any of your data in their AI models. 2. Your employer can pay you and your colleagues to use your collective data, at a price that you all agree on. You and your coworkers could then decide how to split the proceeds amongst yourselves individually."

[13] Specifically, for each policy, we ask: "How would you feel if this policy were in place where you work?" (Positive, I would like it / Neutral / Negative, I would dislike it).

Three key features differentiate our setting from a standard employment relationship:

1. **Two periods**: today's data train tomorrow's AI, creating dynamic effects of knowledge contribution.

2. **Non-contractible knowledge contributions**: the firm does not yet know what skilled workers do until it surveils them and codifies their tacit knowledge, so it cannot write contracts specifying exactly what knowledge workers should provide.

3. **Limited commitment**: workers' expertise walks with them in the status quo—firms cannot guarantee lifetime employment, and workers cannot commit themselves indefinitely. Everything therefore happens under the shadow of renegotiation.

Our model formalizes a dynamic hold-up problem created by surveillance-enabled AI. In period 1, workers reveal tacit know-how while doing their jobs. These records train an AI system that becomes the firm's period-2 outside option. When future bargaining happens under limited commitment, a stronger outside option allows the firm to push down wages. Anticipating this, workers strategically withhold knowledge today, even though withholding is privately costly and reduces current output.

4.2 Model Setup

4.2.1 Players and Timing

The economy consists of one firm and a finite set of workers J. Time is discrete with two periods $t \in \{1, 2\}$, representing the present and the future.

Within each period

1. The firm bargains with each worker j over wage $w_t^j \geqslant 0$ and employment status $I_t^j \in \{0, 1\}$.

2. Each employed worker chooses a non-negative knowledge contribution $k_t^j \in K^j$, where K^j is a finite set.

3. Production occurs and wages are paid.

Between periods, everyone observes the vector of first-period contributions $k_1 \equiv (k_1^1, k_1^2, \ldots, k_1^{|J|})$.

We normalize each worker's output to be equal to their knowledge contribution k_t^j.[14]

[14] Our other assumptions do not rely on the cardinal properties of k_t^j, so if output is some continuous increasing function of k_t^j, we could rescale it so that each k_t^j is equal to the resulting output level.

We denote the vector of time-t contributions $k_t \equiv (k_t^j)_{j \in J}$, and similarly the vector of time-t wages $w_t \equiv (w_t^j)_{j \in J}$. Given some dataset $D \subseteq J$, we use k_t^D to denote the vector that is identical to k_t except that elements corresponding to $j \notin D$ are equal to 0.

We model AI quality using a function $\alpha : \mathbb{R}^J_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$. Given some time-1 contributions k_1 and some dataset D, the firm develops an AI system of quality $\alpha(k_1^D)$. Intuitively, this captures the productivity of an AI system that can augment unskilled workers or even replace workers entirely.

At time 2, for each worker j, the firm with dataset D gets output

$$\max \left\{ I_2^j k_2^j, \alpha(k_1^D) \right\}. \tag{1}$$

The no-surveillance case corresponds to $D = \varnothing$. The functional form of (1) means that the AI automates the role, instead of augmenting the worker, because the gains from hiring worker j arise only when worker j's output exceeds what the AI could separately achieve.

We assume that α is monotone increasing in knowledge inputs: that is, for all $\underline{k}_1 \leqslant \overline{k}_1$, we have $\alpha(\underline{k}_1) \leqslant \alpha(\overline{k}_1)$. We assume that knowledge contributions are substitute inputs; that is, for all $\underline{k}_1^j \leqslant \overline{k}_1^j$ and all $\underline{k}_1^{-j} \leqslant \overline{k}_1^{-j}$, we have

$$\alpha(\overline{k}_1^j, \underline{k}_1^{-j}) - \alpha(\underline{k}_1^j, \underline{k}_1^{-j}) \geqslant \alpha(\overline{k}_1^j, \overline{k}_1^{-j}) - \alpha(\underline{k}_1^j, \overline{k}_1^{-j}). \tag{2}$$

Here are some examples of α that are permitted by our assumptions:

1. Linear returns $\alpha(k_1) = \sum_j \beta_j k_1^j$ for constants $\beta_j \geqslant 0$,
2. Replicating the skills of the best worker $\alpha(k_1) = \beta \max_j k_1^j$ for constant $\beta \geqslant 0$,
3. Decreasing returns to the sum of contributions $\alpha(k_1) = h\big(\sum_j k_1^j\big)$ where $h : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is increasing and concave.

4.2.2 Worker Characteristics

When worker j contributes knowledge k_t^j, they incur private cost $c^j(k_t^j)$. We assume that the cost function $c^j : K^j \to \mathbb{R}_{\geqslant 0}$ has a unique minimizer $\theta^j > 0$, which attains $c^j(\theta^j) = 0$. We interpret θ^j as the worker's skill; this is the level of knowledge that they would naturally reveal when monitored, absent deliberate efforts to withhold knowledge or to record it. In general, the worker may withhold knowledge at a cost ($k_t^j < \theta^j$) or contribute more at a cost ($k_t^j > \theta^j$).

To ease exposition, we assume for now that the worker may only withhold knowledge; that is, $\theta^j = \max K^j$. Under this assumption, the costs capture strategic behavior such as sandbagging, evading surveillance, contaminating data, or degrading documentation. We will later examine how the results change with costly contributions.

4.2.3 Payoffs

We assume that workers' payoffs are additive across time, linear in wage and withholding costs, and that their outside option yields zero payoff. Thus, worker j's utility is

$$U^j = I_1^j \left(w_1^j - c^j(k_1^j) \right) + \psi \, I_2^j \left(w_2^j - c^j(k_2^j) \right), \tag{3}$$

where $\psi > 0$ is the weight on the future. We allow for the case that $\psi > 1$; one can interpret this as meaning that the stakes from expropriation are so large as to outweigh time discounting.

We assume that the firm's payoff is additive across time and across workers, and linear in output and wages. That is, the firm's utility is

$$\Pi = \sum_{j \in J} \left[I_1^j \left(k_1^j - w_1^j \right) + \psi \, \left(\max \left\{ I_2^j k_2^j, \alpha(k_1^D) \right\} - I_2^j w_2^j \right) \right]. \tag{4}$$

4.2.4 Solution concept

Observe that upon being hired at time 2, the unique best response for the worker is to set $k_2^j = \theta^j$, at cost $c^j(\theta^j) = 0$, and we will assume that they so do. For simplicity, we will break ties in firm-worker bargaining in favor of hiring.

An **assessment** in our model is a tuple consisting of:

1. Time-1 wages w_1.
2. Time-1 employment $\overline{J}_1$.
3. Time-1 knowledge contributions $k_1 \in \prod_{j \in J} K^j$.
4. Time-2 wages $\omega_2 : \prod_{j \in J} K^j \to \mathbb{R}^J_{\geqslant 0}$.
5. Time-2 employment $\mathcal{J}_2 : \prod_{j \in J} K^j \to 2^J$.

Note that we have specified contributions k_1^j for workers $j \notin \overline{J}_1$; these are the contributions those workers would make if (counterfactually) they were hired.

We assume that in each period, firms and workers engage in Nash-in-Nash bargaining, with exogenous bargaining weight $\gamma \in [0,1]$ for each worker. In practice, many workplaces bargain bilaterally at the worker level (or via managers), while firm profits depend on the entire workforce. Nash-in-Nash is a tractable way to capture simultaneous bilateral bargaining while accounting for cross-worker spillovers from AI. The exogenous weight γ should be read as reduced-form bargaining strength, reflecting regulations or labor market conditions.

An assessment is an **equilibrium** if:

1. Time-1 wages w_1 and employment $\overline{J}_1$ are a Nash-in-Nash equilibrium, with the disagreement point for the worker j consisting of not being hired and not being paid.

2. For each worker $j \in J$, their contribution k_1^j maximizes their continuation payoff when the other workers contribute according to $k_1^{\overline{J}_1}$.[15]

3. For each $k_1' \in \prod_{j \in J} K^j$, wages $\omega_2(k_1')$ and employment $\mathcal{J}_2(k_1')$ are a Nash-in-Nash equilibrium.

Our models differ in the firm's dataset in the event of agreement and disagreement; we describe each in the subsections that follow.

4.3 Results

4.3.1 No surveillance

Without surveillance, the firm's dataset is always $D = \varnothing$, and the disagreement point has worker j not hired and not paid.

Our model is designed to isolate the dynamic effects of expropriation by AI. Without surveillance, firms and workers are engaging in two identical and separable production stages. Knowledge contributed at time 1 has no effect on the worker's time-2 payoffs, so workers have no incentive to withhold knowledge. We state this formally in the following observation.

Observation 4.1. With no surveillance, every equilibrium has the following features:

1. full employment in both periods,

2. each worker contributes knowledge equal to their skill in both periods, $k_1^j = k_2^j = \theta^j$, and

[15] Implicitly, this means that hired workers do not observe who else is hired when deciding their own contribution; so they best-respond to the *equilibrium* hired set $\overline{J}_1$. Notice also that workers not hired at time-1 have 'passive beliefs'. That is, if (off-path) they are hired, they believe the set of other workers hired is the same.

3. wages $w_1^j = \gamma\theta^j$ and $w_2^j = \gamma\left\lfloor\theta^j - \alpha(0)\right\rfloor_+$, where γ is the worker's Nash bargaining weight and $\lfloor x\rfloor_+$ denotes $\max\{x, 0\}$.

At time 2, the firm's disagreement payoff for each position is $\alpha(0)$, the output of an AI system trained on no data; when $\alpha(0) = 0$, wages are $\gamma\theta^j$ in both periods.

This equilibrium is a useful benchmark, because workers withhold no knowledge and incur no withholding costs. This equilibrium only falls short of the first-best because it achieves none of the productivity gains from training AI on workers' data.

Even under the alternative policies that follow, there will be equilibria with full employment in both periods, because the worker's outside option yields zero payoff. For ease of exposition, we will focus on these equilibria, so that the key policy-relevant comparisons we focus on are: total surplus, worker wages, and firm profits.

4.3.2 Firm-owned AI

With firm-owned AI, the firm's dataset is always $\overline{J}_1$, and the disagreement point has worker j not hired and not paid. Thus, even if no agreement is reached with worker j at time-2, the firm still produces $\alpha(k_1)$ using last period's data.

Suppose the workers naïvely contributed full knowledge at time-1. Holding fixed the workers' time-1 contributions k_1, firm-owned AI raises the output resulting from agreement between the firm and worker j at time-2, from $\max\left\{\theta^j, \alpha(0)\right\}$ to $\max\left\{\theta^j, \alpha\left(k_1^{\overline{J}_1}\right)\right\}$. But it also raises the firm's disagreement payoff, from $\alpha(0)$ to $\alpha\left(k_1^{\overline{J}_1}\right)$. Thus, Nash-in-Nash bargaining results in a fall in the worker's wage, from $\gamma\left\lfloor\theta^j - \alpha(0)\right\rfloor_+$ to $\gamma\left\lfloor\theta^j - \alpha\left(k_1^{\overline{J}_1}\right)\right\rfloor_+$. Thus, ignoring equilibrium effects, firm-owned AI increases time-2 output and the firm's time-2 profits.

Next we study what happens in equilibrium with firm-owned AI. The time-1 contribution game has a useful structure: Even though workers are harmed by raising k_1^j (it strengthens the AI they face tomorrow), the substitutes property implies the marginal harm from revealing more is smaller when others already revealed a lot. Thus, workers' time-1 contributions are strategic complements. Intuitively, if your colleagues have already "taught the AI most of what it needs," your own contribution does little incremental damage to your future wage but still saves you withholding effort $c^j(\cdot)$, so best responses are weakly increasing in others' contributions.

To ensure that our results are not vacuous, we start with a simple sufficient condition for the existence of equilibrium. We require that (essentially) withholding costs are not so severe that they

outweigh a single-worker's time-1 output. Formally, let B^j be the best-response contributions of worker j when all other workers contribute 0, that is

$$B^j \equiv \arg\max_{k_1^j} \left\{ w_1^j - c^j\left(k_1^j\right) + \psi\gamma \left\lfloor \theta^j - \alpha\left(k_1^j, 0\right)\right\rfloor_+ \right\}. \tag{5}$$

We assume that

$$\inf\left\{ k_1^j - c^j\left(k_1^j\right) : k_1^j \in B^j \right\} \geqslant 0. \tag{6}$$

We now state a result that guarantees the existence of full-employment equilibrium, and implies a (weak) reduction in time-2 wages compared to the no-surveillance case.

Theorem 4.2. With firm-owned AI, under assumption (6), there exists an equilibrium with:

1. Full employment in both periods $J = \overline{J}_1 = \mathcal{J}_2(\tilde{k}_1)$ for all $\tilde{k}_1$,
2. Time-2 wages $\omega_2^j(\tilde{k}_1) = \gamma \left\lfloor \theta^j - \alpha(\tilde{k}_1)\right\rfloor_+$ for all $\tilde{k}_1$ and all j.

Proof. We have restricted attention to equilibria with full time-2 knowledge contributions, and in every such equilibrium, Nash-in-Nash bargaining at time 2 implies that $\omega_2^j(\tilde{k}_1) = \gamma \left\lfloor \theta^j - \alpha(\tilde{k}_1)\right\rfloor_+$ for all $\tilde{k}_1$ and all j.

Given some time-1 hired set $\overline{J}_1$ and wages w_1, it follows that worker j chooses k_1^j to maximize the utility function

$$w_1^j - c^j(k_1^j) + \psi\gamma \left\lfloor \theta^j - \alpha\left(k_1^{\overline{J}_1 \cup \{j\}}\right)\right\rfloor_+. \tag{7}$$

In order to show existence, we will establish that the simultaneous choice of k_1^j by the workers to maximize (7) is a supermodular game, in the sense of Milgrom and Roberts (1990). To do so, we first prove a technical lemma.

Lemma 4.3. Let X and Y be partially ordered sets. Suppose $f : X \times Y \to \mathbb{R}$ is monotone non-increasing[16] and has increasing differences. Suppose $g : \mathbb{R} \to \mathbb{R}$ is non-decreasing and convex. Then $g \circ f : X \times Y \to \mathbb{R}$ has increasing differences.

We now prove Lemma 4.3. The function f has increasing differences, so for all $\underline{x} \leqslant \overline{x}$ and $\underline{y} \leqslant \overline{y}$, we have

$$f(\underline{x}, \overline{y}) - f(\overline{x}, \overline{y}) \leqslant f(\underline{x}, \underline{y}) - f(\overline{x}, \underline{y}). \tag{8}$$

[16] That is, for $\underline{x} \leqslant \overline{x}$ and $\underline{y} \leqslant \overline{y}$, we have $f(\underline{x}, \underline{y}) \geqslant f(\overline{x}, \overline{y})$.

By f monotone non-increasing, we have

$$\Phi \equiv f(\underline{x}, \overline{y}) - f(\overline{x}, \overline{y}) \geqslant 0, \tag{9}$$

$$f(\overline{x}, \overline{y}) \leqslant f(\overline{x}, \underline{y}). \tag{10}$$

It follows that

$$\begin{aligned} g(f(\underline{x}, \overline{y})) - g(f(\overline{x}, \overline{y})) &= g(f(\overline{x}, \overline{y}) + \Phi) - g(f(\overline{x}, \overline{y})) \\ &\leqslant g(f(\overline{x}, \underline{y}) + \Phi) - g(f(\overline{x}, \underline{y})) \text{ by (9), (10) and } g \text{ convex} \\ &\leqslant g(f(\underline{x}, \underline{y})) - g(f(\overline{x}, \underline{y})) \text{ by (8) and } g \text{ non-decreasing.} \end{aligned} \tag{11}$$

Thus, $g \circ f$ has increasing differences. This completes the proof of Lemma 4.3.

Lemma 4.4. The simultaneous choice of k_1^j by the workers to maximize (7) is a supermodular game.

Most of the requirements for a supermodular game follow by inspection. The only non-trivial part is to show that for each worker j, the utility function

$$w_1^j - c^j(k_1^j) + \psi\gamma \left\lfloor \theta^j - \alpha \left(k_1^{\overline{J}_1 \cup \{j\}} \right) \right\rfloor_+ \tag{12}$$

has increasing differences in (k_1^j, k_1^{-j}). Observe that $\theta^j - \alpha \left(k_1^{\overline{J}_1 \cup \{j\}} \right)$ has increasing differences in (k_1^j, k_1^{-j}) by the assumption that worker contributions are substitutes, and it is monotone non-increasing. It follows that

$$\psi\gamma \left\lfloor \theta^j - \alpha \left(k_1^{\overline{J}_1 \cup \{j\}} \right) \right\rfloor_+ \tag{13}$$

has increasing differences by Lemma 4.3. Moreover, $w_1^j - c^j(k_1^j)$ has increasing differences trivially, because it does not depend on k_1^{-j}. The set of functions with increasing differences is closed under addition, so it follows that (12) has increasing differences in (k_1^j, k_1^{-j}). This completes the proof of Lemma 4.4.

Fixing time-1 wages w_1 and employment $\overline{J}_1$, there exists a contribution profile k_1 that satisfies the requirements of equilibrium, by Lemma 4.4 and Theorem 5 of Milgrom and Roberts (1990).

We now guess and verify that full employment at both periods, that is, $J = \overline{J}_1 = \mathcal{J}_2(\tilde{k}_1)$ for all $\tilde{k}_1$, is part of an equilibrium. This follows straightforwardly for time 2 because ω_2^j derived above gives the firm a non-negative payoff from hiring each worker.

We now consider time 1. Hiring worker j at wage w_1^j results in firm payoff

$$k_1^j - w_1^j + \sum_{l \neq j} (k_1^l - w_1^l) + \psi \sum_l \left(\alpha(k_1^J) + (1-\gamma) \left\lfloor \theta^l - \alpha(k_1^J) \right\rfloor_+ \right) \tag{14}$$

and worker payoff

$$w_1^j - c^j(k_1^j) + \psi\gamma \left\lfloor \theta^j - \alpha(k_1^J) \right\rfloor_+ . \tag{15}$$

Not hiring worker j results in firm payoff

$$\sum_{l \neq j} (k_1^l - w_1^l) + \psi \sum_l \left(\alpha(k_1^{J\setminus\{j\}}) + (1-\gamma) \left\lfloor \theta^l - \alpha(k_1^{J\setminus\{j\}}) \right\rfloor_+ \right), \tag{16}$$

and worker payoff

$$\psi\gamma \left\lfloor \theta^j - \alpha(k_1^{J\setminus\{j\}}) \right\rfloor_+ \tag{17}$$

Thus, compared to the disagreement point, hiring worker j at time 1 increases the pairwise surplus by

$$\begin{aligned} & k_1^j - c^j(k_1^j) + \psi \left(\max\left\{\theta^j, \alpha(k_1^J)\right\} - \max\left\{\theta^j, \alpha(k_1^{J\setminus\{j\}})\right\} \right) \\ & \quad + \psi \sum_{l \neq j} \left(\alpha(k_1^J) + (1-\gamma) \left\lfloor \theta^l - \alpha(k_1^J) \right\rfloor_+ - \alpha(k_1^{J\setminus\{j\}}) - (1-\gamma) \left\lfloor \theta^l - \alpha(k_1^{J\setminus\{j\}}) \right\rfloor_+ \right). \end{aligned} \tag{18}$$

Next we show that

$$k_1^j - c^j(k_1^j) \geqslant 0. \tag{19}$$

By hypothesis, the contribution profile k_1 is a pure-strategy Nash equilibrium of the supermodular game considered in Lemma 4.4. It follows that

$$k_1^j \in \arg\max_{\hat{k}_1^j} \left\{ w_1^j - c^j\left(\hat{k}_1^j\right) + \psi\gamma \left\lfloor \theta^j - \alpha\left(\hat{k}_1^j, k_1^{-j}\right) \right\rfloor_+ \right\}. \tag{20}$$

Let $\underline{k}_1^j$ be the smallest element of the right-hand side of (5). By Topkis' theorem (Milgrom and Shannon, 1994), we have that $\underline{k}_1^j \leqslant k_1^j$. Then by (20) and α non-decreasing we have that $c^j(\underline{k}_1^j) \geqslant$

$c^j(k_1^j)$. By the preceding inequalities and (6), we have

$$0 \leqslant \inf\left\{\hat{k}_1^j - c^j\left(\hat{k}_1^j\right) : \hat{k}_1^j \in B^j\right\} \leqslant \underline{k}_1^j - c^j\left(\underline{k}_1^j\right) \leqslant k_1^j - c^j\left(k_1^j\right). \tag{21}$$

We have proved (19). By (19) and α monotone non-decreasing, it follows that (18) is non-negative. We have proved that full employment at time 1 is consistent with Nash-in-Nash bargaining, which completes the proof. □

We henceforth restrict attention to equilibria of the form characterized in Theorem 4.2.

Observation 4.5. With firm-owned AI, each worker's time-2 wage is lower than the no-surveillance baseline, and strictly lower if the time-1 knowledge contributions strictly improve AI quality and the baseline wage is positive, that is if $\alpha(k_1) > \alpha(0)$ and $\alpha(0) < \theta^j$.

Observation 4.6. With firm-owned AI, workers withhold knowledge at time 1 compared to the no-surveillance baseline, that is for every worker j, we have $k_1^j \leqslant \theta^j$.

Proof. With no surveillance, each worker's time-1 knowledge contribution maximizes $k_1^j - c^j(k_1^j)$, which has a unique maximizer $k_1^j = \theta^j$. With firm-owned AI, each worker's time-1 knowledge contribution maximizes (12). The observation follows by Topkis's theorem. □

Observation 4.7. In equilibrium, knowledge withholding increases worker surplus and decreases firm surplus, compared to what would happen if each worker naïvely contributed $k_1^j = \theta^j$.

Proof. Let us define

$$u^j(k_1) \equiv w_1^j - c^j(k_1^j) + \psi\gamma\left[\theta^j - \alpha\left(k_1\right)\right]_+. \tag{22}$$

Let k_1 denote an equilibrium time-1 contribution profile. We have

$$u^j(k_1^j, k_1^{-j}) \geqslant u^j(\theta^j, k_1^{-j}) \geqslant u^j(\theta^j, \theta^{-j}), \tag{23}$$

where the first inequality is by worker best-response, and the second inequality is by u^j decreasing in the other workers' time-1 contributions. We have proved that knowledge withholding increases worker surplus. Since withholding is costly and reduces time-2 output, it reduces total surplus. It follows that it reduces firm surplus. □

Under what conditions do workers withhold more knowledge in equilibrium? We now state some comparative statics results, restricting attention to equilibria of the form characterized by Theorem 4.2.

Our assumptions allow for the possibility of multiple equilibria, which can complicate comparative statics. But the workers' time-1 contributions are strategic complements, by Lemma 4.4. Thus, holding fixed the primitives, there exists a highest equilibrium, in the sense that each worker contributes weakly more at time 1 than they do in any other equilibrium (Milgrom and Roberts, 1990, Theorem 5). And there also exists a lowest equilibrium, in the sense that each worker contributes weakly less than they do in any other equilibrium. We will keep track of the highest and lowest equilibria as the primitives change.

First, we show that workers withhold less (contribute more) when withholding has a higher marginal cost.

Theorem 4.8. Consider two profiles of cost functions, $(c^j)_{j\in J}$ and $(\tilde{c}^j)_{j\in J}$, where for all workers j and contributions $k_1^j \leqslant k_1^{j'}$, we have

$$c^j(k_1^j) - c^j(k_1^{j'}) \leqslant \tilde{c}^j(k_1^j) - \tilde{c}^j(k_1^{j'}). \tag{24}$$

Holding all other primitives fixed, let $\overline{k}_1$ and $\underline{k}_1$ denote the highest and lowest equilibrium time-1 contributions under $(c^j)_{j\in J}$, and similarly let $\overline{\tilde{k}}_1$ and $\underline{\tilde{k}}_1$ denote the highest and lowest equilibrium time-1 contributions under $(\tilde{c}^j)_{j\in J}$. We have $\overline{k}_1 \leqslant \overline{\tilde{k}}_1$ and $\underline{k}_1 \leqslant \underline{\tilde{k}}_1$.

Proof. By Lemma 4.4, the worker payoffs (7) from time-1 contributions induced by $(c^j)_{j\in J}$ and $(\tilde{c}^j)_{j\in J}$ define a pair of supermodular games indexed by parameter τ, with $\tau = 0$ corresponding to $(c^j)_{j\in J}$ and $\tau = 1$ corresponding to $(\tilde{c}^j)_{j\in J}$. Each worker j's utility has increasing differences in (k_1^j, τ) by (24). Thus, the result follows by Theorem 6 of Milgrom and Roberts (1990). □

Next we show that equilibrium contributions rise when workers are better substitutes for each other. To state this formally, we slightly extend the model: Suppose that each worker occupies a distinct role, and let the AI quality in role j be denote $\alpha^j(k_1)$. All the results stated so far extend to this case, with essentially the same proofs.

Suppose we found a better way to use other workers' data to automate worker j's role. Intuitively, this would raise the AI quality in worker j's role for a given profile of contributions. And it would

lower the marginal returns to AI quality from raising worker j's contribution. We now formally show that such a change increases equilibrium contributions.

Given two AI technologies, $\left(\alpha^j\right)_{j\in J}$ and $\left(\tilde{\alpha}^j\right)_{j\in J}$, we write that $\left(\tilde{\alpha}^j\right)_{j\in J}$ is **at least as good as** $\left(\alpha^j\right)_{j\in J}$ if for each role j, we have

$$\alpha^j(k_1) \leqslant \tilde{\alpha}^j(k_1) \text{ for all } k_1. \tag{25}$$

Given two AI technologies, $\left(\alpha^j\right)_{j\in J}$ and $\left(\tilde{\alpha}^j\right)_{j\in J}$, we write that $\left(\tilde{\alpha}^j\right)_{j\in J}$ has **stronger substitution than** $\left(\alpha^j\right)_{j\in J}$ if for each role j, we have

$$\alpha^j(\hat{k}_1^j, k_1^{-j}) - \alpha^j(k_1^j, k_1^{-j}) \geqslant \tilde{\alpha}^j(\hat{k}_1^j, k_1^{-j}) - \tilde{\alpha}^j(k_1^j, k_1^{-j}) \text{ for all } k_1^j \leqslant \hat{k}_1^j \text{ and all } k_1^{-j}. \tag{26}$$

Theorem 4.9. Suppose that $\left(\tilde{\alpha}^j\right)_{j\in J}$ is at least as good as $\left(\alpha^j\right)_{j\in J}$ and that $\left(\tilde{\alpha}^j\right)_{j\in J}$ has stronger substitution than $\left(\alpha^j\right)_{j\in J}$. Holding all other primitives fixed, let $\overline{k}_1$ and $\underline{k}_1$ denote the highest and lowest equilibrium time-1 contributions under $\left(\alpha^j\right)_{j\in J}$, and similarly let $\overline{\tilde{k}}_1$ and $\underline{\tilde{k}}_1$ denote the highest and lowest equilibrium time-1 contributions under $\left(\tilde{\alpha}^j\right)_{j\in J}$. We have $\overline{k}_1 \leqslant \overline{\tilde{k}}_1$ and $\underline{k}_1 \leqslant \underline{\tilde{k}}_1$.

Moreover, worker wages at time-2 are lower under technology $\tilde{\alpha}$ and contribution profiles $\overline{\tilde{k}}_1$ and $\underline{\tilde{k}}_1$, compared to technology α and contribution profiles $\overline{k}_1$ and $\underline{k}_1$ respectively.

The proof is in Section B.1.

4.4 Alternative Policies

4.4.1 Individual data ownership

Under individual-owned AI, the firm's dataset D is equal to the set of workers hired at *both* time-1 and time-2. Thus, the disagreement point for bargaining at time-2 between the firm and worker j involves the worker not being hired, not being paid, and their data being omitted from the dataset.

Since we are breaking ties in favor of employment and the pairwise surplus from employing an additional worker is at least zero, every equilibrium with individual-owned AI involves full employment at time 2. In such an equilibrium, the pairwise surplus from employing worker j at time-2 is

$$\lambda_j(k_1, \overline{J}_1) \equiv \max\left\{\theta^j, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \alpha\left(k_1^{\overline{J}_1\setminus\{j\}}\right) + \sum_{l\neq j}\left(\max\left\{\theta^l, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \max\left\{\theta^l, \alpha\left(k_1^{\overline{J}_1\setminus\{j\}}\right)\right\}\right) \tag{27}$$

Observe that $\lambda_j(k_1, \overline{J}_1)$ is non-decreasing in k_1^j, by α monotone non-decreasing. By Nash-in-Nash bargaining, worker j's equilibrium wage is $\gamma\lambda_j(k_1, \overline{J}_1)$. At time 1, worker j chooses k_1^j to maximize

$$w_1^j - c^j(k_1^j) + \psi\gamma\lambda_j(k_1, \overline{J}_1), \tag{28}$$

and by c^j non-increasing and λ_j non-decreasing in k_1^j, it follows that it is a best response to contribute $k_1^j = \theta^j$.

At time 1, however, the pairwise surplus from employing worker j need not be positive. That surplus includes the effect of worker j's employment on the firm's time-2 wage bill: worker j's data enters the firm's disagreement point with every other worker l, and thereby changes every other worker's wage $\gamma\lambda_l$. When worker j's data pushes AI quality across the other workers' skill levels, the cross-position terms in (27) switch on, and the wage bill can rise by more than output. As the following example shows, individual ownership can then be inconsistent with full employment at time 1—and, as a consequence, individual ownership can yield strictly lower total surplus than firm ownership, and even than the no-AI benchmark.

Example 1. There are three workers with $\theta^1 = \theta^2 = \theta^3 = 10$ and $K^j = \{10\}$, so that $c^j \equiv 0$: withholding is impossible. Let $\gamma = 1$ and $\psi = 11$, and let AI quality depend on the number of workers whose data the firm holds: quality 0, 6, 10, or 13 for datasets of size 0, 1, 2, or 3. This α is monotone with weakly decreasing increments, so (2) holds. If all three workers are employed at time 1, then by (27) each worker's time-2 wage is $(13-10)+2\times(13-10) = 9$: removing any one worker's data lowers AI quality from 13 to 10, destroying the surplus $13-10$ on all three positions. The firm's time-2 profit is $3\times 13 - 3\times 9 = 12$. If instead some worker j is not employed at time 1, then time-2 output falls to $3\times 10 = 30$, but the two employed workers' wages fall to $10-6 = 4$ each and worker j's wage to 0, so the firm's time-2 profit rises to $30-8 = 22$. Employing worker j at time 1 thus raises time-2 output by 9 but raises the *other* workers' wages by 10: the joint continuation payoff of the firm and worker j falls by 1, and the time-1 pairwise surplus is $10+\psi\times(-1) < 0$. Full employment at time 1 is therefore inconsistent with Nash-in-Nash bargaining. Indeed, every equilibrium employs exactly two workers at time 1 (the analogous computation gives pairwise surplus $10+11\times 12 = 142$ for a first hire and $10+11\times 10 = 120$ for a second hire). Total surplus is $20+\psi\times 30 = 350$, strictly below the no-AI benchmark, $30+\psi\times 30 = 360$, and strictly below firm ownership, which sustains full employment (Theorem 4.2) and attains the first best, $30+\psi\times 39 = 459$, because withholding is impossible.

The failure in this example is a threshold effect: a worker's data affects the other workers' wages only through terms of the form $\max\{\theta^m, \alpha(\cdot)\}$, so the wage bill responds discontinuously when the dataset pushes AI quality across a worker's skill level. The following condition rules this out.

Definition 4.10 (Condition C: no threshold crossing)**.** Condition C holds if, for every contribution profile $k_1 \in \prod_{j \in J} K^j$ with $k_1^j \geqslant \theta^j$ for all j, and every worker m: either $\theta^m \geqslant \alpha\left(k_1^J\right)$, or $\theta^m \leqslant \alpha\left(k_1^{J\setminus\{l,l'\}}\right)$ for all $l, l' \in J$.

Condition C says that whenever the AI trained on everyone's data strictly outperforms worker m, the AI trained without any *two* workers' data still weakly outperforms worker m. Datasets missing two workers are the relevant ones because the time-1 pairwise surplus nests two disagreements: the firm's time-2 disagreement point with worker l, evaluated in the branch where worker j is not employed at time 1, involves the dataset missing both j and l. The restriction to profiles with $k_1^j \geqslant \theta^j$ is what equilibrium play delivers, because under individual ownership no worker withholds. In the present model, where $\theta^j = \max K^j$, the only such profile is $k_1 = \theta$; the general form of the condition matters for the extension to costly contributions below.

Proposition 4.11. Under individual-owned AI, if Condition C holds, then there is an equilibrium with full employment in both periods and with $k_1^j = k_2^j = \theta^j$.

The proof is in Section B.2.

Henceforth we assume Condition C and restrict attention to the equilibrium described in Proposition 4.11. Notice that, compared to firm-owned AI, workers contribute weakly more knowledge at time 1, which raises output directly at time 1 and at time 2 indirectly, by improving AI quality.

Observation 4.12. Total surplus is higher under individual ownership than under firm ownership.

However, individual ownership does not guarantee that workers are better off, compared to the no-AI baseline. The intuition for this is that Nash-in-Nash bargaining compensates workers based on the marginal contribution of their data, and under substitutes the total contribution exceeds the sum of the marginal contributions. That is,

$$\alpha(k_1^J) - \alpha(0) \geqslant \sum_j \left(\alpha(k_1^J) - \alpha(k_1^{J\setminus\{j\}})\right). \tag{29}$$

To see this, observe that we can number the workers arbitrarily from 1 to $|J|$, and rewrite the left-hand side of (29) as a telescoping sum

$$\alpha(k_1^J) - \alpha(0) = \sum_{l=1}^{|J|} \left(\alpha(k_1^{\{j:j \leqslant l\}}) - \alpha(k_1^{\{j:j \leqslant l-1\}}) \right) \geqslant \sum_j \left(\alpha(k_1^J) - \alpha(k_1^{J \setminus \{j\}}) \right), \tag{30}$$

where the inequality follows by the substitutes assumption.

Individual ownership generates a competition externality: Each worker accounts for the (non-negative) effect of their time-1 contribution on their time-2 wage, but does not account for how their time-1 contribution decreases other workers' time-2 wage. Thus, individual ownership does not ensure that workers receive a substantial share of the rents from AI. Workers' time-2 wages are lower when their data are more substitutable for each other, as we now state formally. For this result, we extend the substitution comparison to cross-worker effects, requiring that removing any one worker's data lowers the AI quality in every role by less under the new technology.

Theorem 4.13. Suppose that for all workers j and all roles l,

$$\alpha^l(\theta) - \alpha^l\left(0, \theta^{-j}\right) \geqslant \tilde{\alpha}^l(\theta) - \tilde{\alpha}^l\left(0, \theta^{-j}\right), \tag{31}$$

and also that

$$\tilde{\alpha}^j(\theta) = \alpha^j(\theta) \text{ for all } j. \tag{32}$$

Then, in the equilibrium of Proposition 4.11, each worker's time-2 wage is weakly lower under $\left(\tilde{\alpha}^j\right)_{j \in J}$ than under $\left(\alpha^j\right)_{j \in J}$.

For $l = j$, condition (31) is an instance of (26), evaluated at the full-contribution profile. The proof is in Section B.3.

Theorem 4.13 concerns time-2 wages; notably, stronger substitution need not lower workers' *total* equilibrium payoffs, because time-1 wages can move in the opposite direction. When time-1 wages are determined by Nash-in-Nash bargaining that accounts for continuation payoffs (as in the proof of Theorem 4.2), the surplus from hiring worker j at time 1 includes the effect of j's data on the firm's time-2 bargaining position against the *other* workers: employing j lowers colleagues' time-2 wages, and worker j captures a share γ of the firm's resulting saving. Stronger substitution makes each worker's data more damaging to colleagues' wages, and so raises this component of time-1 pay. To illustrate, suppose there are two workers with $\theta^1 = \theta^2 = 5$ and $\psi = \gamma = 1$, and

consider a technology such that AI quality in either role is 0, 6, or 8 when the firm holds zero, one, or both workers' data. Compare this to an alternative technology with corresponding qualities 1, 7, and 8: the alternative satisfies (31) and (32), and its higher data-free quality can be interpreted as improved pretrained capability. Moving to the alternative technology, each worker's time-2 wage falls from 4 to 2, in accordance with Theorem 4.13, but each worker's time-1 wage rises from 8 to 11, so each worker's total payoff *rises* from 12 to 13. In this sense, individual data ownership pays workers partly for undermining their colleagues, and this component of pay grows as workers' data become better substitutes.

In some cases, individual ownership can harm workers compared to the no-AI baseline, even when workers have full Nash bargaining power. We now state this formally.

Theorem 4.14. Suppose that:

1. There are at least three workers,
2. workers have identical maximum contributions, with $\theta^j = \theta^{j'}$ for all j and j',
3. contributions are perfect substitutes, that is $\alpha(k_1) = f(\max_j\{k_1^j\})$ for some increasing function f with $f(0) = 0$.

Then for any Nash bargaining parameter $\gamma \in (0, 1]$, the full-contribution full-employment equilibrium under individual data ownership is strictly worse for each worker than the full-contribution full-employment equilibrium under no surveillance.

The proof is in Section B.4.

To summarize, individual data ownership restores efficiency, because it ensures that each worker has no incentive to withhold knowledge at time-1. But individual data ownership does not guarantee that workers share in the efficiency gains from AI. By Theorem 4.14, under some conditions workers can be strictly worse off under individual data ownership, compared to the no-surveillance case. Under those conditions, firms benefit from AI not only because it raises time-2 output, but also because it suppresses workers' time-2 wages.

4.4.2 Collective data ownership

So far we have considered regimes in which data rights are tied to individual workers. Suppose instead that the workers pool their work data in a **data trust**. The trust controls the pooled dataset

and bargains with the firm over its use; employment and wages remain individually negotiated, via Nash-in-Nash bargaining, exactly as before. This institution matches the collective policy in our survey (Section 3.6): workers jointly decide whether their collective work data can be used for AI training, and the employer can pay for its use, but there is no collective bargaining over wages.

Formally, the members of the trust are the workers employed at time 1 (those whose data exists), and the trust controls the dataset $D = \overline{J}_1$. At time 2:

1. The firm and the trust bargain over the use of the pooled dataset, with Nash bargaining weight $\eta \in [0,1]$ for the trust. In the event of disagreement, the firm's dataset is $D = \varnothing$, and wages are negotiated in that no-data environment: the data-use agreement is the senior arrangement, and the wage bargains are negotiated in its shadow.

2. Given the outcome of the data bargain, the firm bargains bilaterally with each worker over employment and wages, with weight γ, as before. In the event of disagreement with an individual worker, that worker is not hired and not paid, but their data remains in the pool, available to the firm under the data-use agreement.

We allow the trust's bargaining weight η to differ from the workers' individual bargaining weight γ.

Because an individual disagreement no longer removes data from the firm's dataset, the pairwise surplus from employing worker l at time 2 is $\max\left\{\theta^l, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \alpha\left(k_1^{\overline{J}_1}\right) = \left\lfloor \theta^l - \alpha\left(k_1^{\overline{J}_1}\right)\right\rfloor_+$, so bilateral wages are $\gamma\left\lfloor \theta^l - \alpha\left(k_1^{\overline{J}_1}\right)\right\rfloor_+$, and full employment at time 2 follows from tie-breaking in favor of hiring. The joint surplus of the data-use agreement is the total output gain that it enables, and the workers' disagreement value is the no-data wage bill $\gamma\sum_{m\in J}\lfloor\theta^m - \alpha(0)\rfloor_+$. Nash bargaining between the firm and the trust therefore gives the worker side a time-2 total of

$$\gamma \sum_{m\in J} \lfloor \theta^m - \alpha(0)\rfloor_+ \;+\; \eta \sum_{m\in J}\left(\max\left\{\theta^m, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \max\left\{\theta^m, \alpha(0)\right\}\right), \tag{33}$$

paid as bilateral wages plus a fee to the trust.

The trust divides the fee among its members. Observe that so long as worker j's time-2 compensation is non-decreasing in k_1^j, it is a best response for worker j to contribute $k_1^j = \theta^j$ at time 1. One division with this property (the **own-gains rule**) gives each member their no-data wage plus a share η of the AI-driven output gain on their own position:

$$\phi^j\left(k_1, \overline{J}_1\right) \equiv \gamma\left\lfloor \theta^j - \alpha(0)\right\rfloor_+ \;+\; \eta\left(\max\left\{\theta^j, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \max\left\{\theta^j, \alpha(0)\right\}\right). \tag{34}$$

Each member's implied fee share is non-negative, and (34) sums to (33). Workers employed at time 2 whose data is not in the pool receive only the bilateral wage.

Remark 4.15. The own-gains rule is one example of a division rule with the properties that the results of this subsection require; it is not essential. The worker side's total (33)—and hence every result that compares totals—does not depend on how the trust divides the fee. Full knowledge contributions require only that each member's compensation be non-decreasing in their own contribution. Full employment at time 1 additionally requires that membership not be a net tax on the member: it suffices that each member receives at least their no-data wage $\gamma \left[\theta^j - \alpha(0)\right]_+$, as under (34). Rules that redistribute strongly across members can violate these properties when workers' skills differ—under an equal split of the fee, for example, a member's compensation can be decreasing in their own contribution—though with symmetric workers, equal splits coincide with (34).

Proposition 4.16. Under the data trust with the own-gains rule, for all $\gamma, \eta \in [0,1]$ and all $\psi > 0$, there is an equilibrium with full employment in both periods and $k_1^j = k_2^j = \theta^j$, in which each worker's time-2 compensation is (34). No analogue of Condition C is required.

The proof is in Section B.5. The contrast with individual ownership is instructive. Under individual ownership, a worker's data is tied to their own employment, so an employment disagreement moves the dataset and hence every other position; a colleague's wage is a marginal-contribution object that can jump by *more* than output when the dataset crosses a skill threshold, producing the hiring freeze of Example 1, and Condition C is needed to rule this out. Under the trust, an individual quit or dismissal no longer moves the dataset, and compensation (34) is a proportional share of own-position output, so the wage bill can never outrun output: the firm keeps a share $1-\eta$ of every output increment, and hiring is always jointly profitable. In the economy of Example 1, the data trust restores full employment, workers each receive 13 (at $\gamma = \eta = 1$), and total surplus is the first best, 459.

Under the data trust, the disagreement point of the data bargain excludes *all* workers' data. This bundles individual data rights so that when one worker contributes, they do not inadvertently strengthen the firm's hand against others. This prevents competition externalities from arising, ensuring that workers share in the gains from AI. Next, we compare time-2 compensation under the two ownership regimes. Under individual ownership, Nash-in-Nash bargaining pays each worker for the *marginal* contribution of their data and, by (29), the sum of marginal contributions falls short

of the total contribution when data are substitutes. The trust instead sells the data as a bundle, so the workers are paid, as a group, for its total contribution. The next observation shows that workers' aggregate time-2 compensation is accordingly higher under the data trust.

Observation 4.17. Fix a technology $\left(\alpha^j\right)_{j\in J}$, and consider the full-contribution full-employment equilibria under individual ownership and under the data trust with $\eta \geqslant \gamma$. Total time-2 payments to workers are weakly higher under the data trust than under individual ownership.

Proof. For any $z \geqslant 0$ we have $\lfloor \theta - z \rfloor_+ - \max\{\theta, z\} = -z$. Applying this identity with $z = \alpha^l(0)$ to each term of (33), written for the role-specific technology, total time-2 payments to workers under the data trust at $\eta = \gamma$ equal

$$\gamma \sum_{l\in J} \left(\max\left\{\theta^l, \alpha^l(\theta)\right\} - \alpha^l(0)\right), \tag{35}$$

and (33) is non-decreasing in η, so it suffices to prove the claim at $\eta = \gamma$. Under individual ownership, total time-2 compensation is $\gamma \sum_j \lambda_j(\theta, J)$, with λ_j as in (27). Grouping the terms of $\sum_j \lambda_j$ by role, we have

$$\sum_j \lambda_j(\theta, J) = \sum_l \left(\max\left\{\theta^l, \alpha^l(\theta)\right\} - \alpha^l\left(0, \theta^{-l}\right) + \sum_{j\neq l} \left(\max\left\{\theta^l, \alpha^l(\theta)\right\} - \max\left\{\theta^l, \alpha^l\left(0, \theta^{-j}\right)\right\}\right)\right). \tag{36}$$

For $x \geqslant y$ we have $\max\{\theta^l, x\} - \max\{\theta^l, y\} \leqslant x - y$, so for each role l,

$$\sum_{j\neq l} \left(\max\left\{\theta^l, \alpha^l(\theta)\right\} - \max\left\{\theta^l, \alpha^l\left(0, \theta^{-j}\right)\right\}\right) \leqslant \sum_{j\neq l} \left(\alpha^l(\theta) - \alpha^l\left(0, \theta^{-j}\right)\right) \leqslant \alpha^l\left(0, \theta^{-l}\right) - \alpha^l(0), \tag{37}$$

where the last inequality holds because, by (29) applied to the function α^l, the sum over *all* $j \in J$ of the marginal contributions $\alpha^l(\theta) - \alpha^l\left(0, \theta^{-j}\right)$ is at most $\alpha^l(\theta) - \alpha^l(0)$, and the $j = l$ term of that sum equals $\alpha^l(\theta) - \alpha^l\left(0, \theta^{-l}\right)$. Substituting (37) into (36) yields

$$\sum_j \lambda_j(\theta, J) \leqslant \sum_l \left(\max\left\{\theta^l, \alpha^l(\theta)\right\} - \alpha^l(0)\right), \tag{38}$$

which completes the proof. □

Combining Observation 4.17 with Theorem 4.13 yields a comparison across technologies. Suppose two technologies are equally good at the two endpoints—with all workers' data, and with no

data—but the marginal contribution of each worker's data is smaller under the second. Moving to the second technology then redistributes the value of the data away from individual workers, and the advantage of the data trust grows.

Corollary 4.18. Suppose that $\left(\alpha^j\right)_{j\in J}$ and $\left(\tilde{\alpha}^j\right)_{j\in J}$ satisfy (31) and (32), and in addition that

$$\tilde{\alpha}^l(0) = \alpha^l(0) \text{ for all } l. \tag{39}$$

Then the difference between total time-2 payments to workers under the data trust and under individual ownership is weakly larger under $\left(\tilde{\alpha}^j\right)_{j\in J}$ than under $\left(\alpha^j\right)_{j\in J}$.

Proof. Total time-2 compensation under the data trust, (33), is identical under the two technologies, by (32) and (39). By Theorem 4.13, each worker's time-2 wage under individual ownership is weakly lower under $\left(\tilde{\alpha}^j\right)_{j\in J}$ than under $\left(\alpha^j\right)_{j\in J}$. □

We close this subsection with two remarks on institutional design. First, the data trust delivers the same outcomes as a conventional union that bargains over employment, wages, and data jointly. Given the dataset, workers' contributions to output are additively separable across positions, so it does not matter whether *labor* is sold individually or collectively; all that matters is that *data* is controlled collectively.

Observation 4.19. Suppose instead that the workers bargain with the firm as a single union with Nash bargaining weight η, jointly over employment, wages, and data use, with a utility equal to the sum of the individual worker utilities; in the event of disagreement, no workers are employed and the firm's dataset is $D = \varnothing$. If $\eta = \gamma$, then total time-2 worker compensation under the union equals total time-2 worker compensation under the data trust, (33).

Proof. The union's Nash payoff at time 2 is η times the joint surplus over the disagreement point, $\eta \sum_{m\in J}\left(\max\left\{\theta^m, \alpha\left(k_1^{\overline{J}_1}\right)\right\} - \alpha(0)\right)$. By the identity $\lfloor\theta - z\rfloor_+ - \max\{\theta, z\} = -z$ applied with $z = \alpha(0)$ to each term of (33), the two expressions coincide when $\eta = \gamma$. □

Second, the two institutions differ in their robustness to the collective's bargaining weight. Under the union, every dollar of worker compensation is bargained at the collective weight η, so a weak collective ($\eta < \gamma$) erodes workers' baseline labor rents along with their data rents. Under the data trust, the collective weight applies only to the AI-driven increment.

Observation 4.20. Under the data trust with the own-gains rule, for every $\eta \in [0,1]$, each member's time-2 compensation (34) is at least their time-2 wage in the no-surveillance benchmark, $\gamma \left[\theta^j - \alpha(0)\right]_+$.

This is immediate from (34), because the second term is non-negative. However weak the trust's bargaining position, its members' time-2 compensation never falls below the no-surveillance benchmark—in contrast to individual ownership, where even full bargaining power cannot prevent every worker's payoff from falling below that benchmark (Theorem 4.14).

Observation 4.20 concerns time-2 compensation, not total payoffs. The trust controls the pooled dataset, but joining the pool is decided in the time-1 employment bargain, which remains bilateral. A worker whose colleagues' data can already replace them ($\alpha\left(0, \theta^{-j}\right) \geqslant \theta^j$) has a poor disagreement position at time 1: if they are never hired, their data never enters the pool, and their time-2 wage $\gamma \left[\theta^j - \alpha\left(0, \theta^{-j}\right)\right]_+$ is zero. Nash-in-Nash bargaining at time 1 then lets the firm capture part of the worker's future data rents through a lower time-1 wage. In the economy of Example 1 with $\gamma = \eta = 1$, each worker's total equilibrium payoff under the trust is $10 + 3\psi$, below the no-AI payoff $10\,(1+\psi)$ for every $\psi > 0$; the comparison remains negative for every $\gamma > 0$ and every $\eta \in [0,1]$. The trust restores full employment, attains the first best, and protects time-2 data rents, but it does not protect time-1 wages. Indeed, because anticipated time-2 compensation is offset one-for-one in the time-1 wage bargain, each worker's equilibrium payoff in this example is non-increasing—and the firm's payoff non-decreasing—in the trust's weight η.

The rents extracted at time 1 do not vanish: they accrue to the firm, alongside the productivity gains from full contributions. This yields an unconditional guarantee on the firm's side of the ledger.

Theorem 4.21. Under the data trust with the own-gains rule, for all $\gamma, \eta \in [0,1]$ and all $\psi > 0$, the firm's payoff in the equilibrium of Proposition 4.16 is weakly greater than its payoff in the no-surveillance equilibrium of Observation 4.1.

The proof is in Section B.6.

4.5 Extension to costly contributions

So far we have assumed that the worker may only withhold knowledge at a cost; that is, $\theta^j = \max K^j$. We now consider the more general model, that permits $\theta^j < \max K^j$. Recall that θ^j is defined as the unique minimizer of the worker's cost function, which attains $c^j(\theta^j) = 0$. Moreover, this is the worker's equilibrium contribution in the no-surveillance baseline. Thus, by permitting $\theta^j < \max K^j$,

we are positing that the worker may contribute strictly more knowledge than that baseline, at some personal cost.

Our analysis of the no-surveillance baseline and of firm-owned AI extends without modification to the case with costly contributions. The results are true as stated, and the proofs do not rely on the assumption that $\theta^j = \max K^j$.[17]

Under individual data ownership, allowing costly contributions changes the analysis technically and substantively. When $\theta^j = \max K^j$, the worker's time-1 best response is to contribute $k_1^j = \theta^j$ regardless of the other workers' contributions, because their time-2 wage is weakly increasing in k_1^j and their time-1 costs are strictly decreasing in k_1^j. By contrast, if $\theta^j < \max K^j$, then the worker's time-1 best response may depend on the other workers' contributions.

Consequently, the technical change to the analysis is that there may not be an equilibrium in pure strategies for the workers.[18] Nonetheless, if we extend the definition of equilibrium to allow workers to mix over time-1 knowledge contributions, then an equilibrium exists. Since each worker's set of possible contributions K^j is finite, and the continuation payoffs resulting from each contribution are well-defined (as in the proof of Theorem 4.2), this follows by the existence of Nash equilibrium for finite games. Under Condition C and a bound on withholding costs, such an equilibrium also has full employment in both periods.

Theorem 4.22. Suppose that Condition C holds, and that for every worker j, we have

$$\inf_{k_1^j \in K^j} \left\{ k_1^j - c^j(k_1^j) \right\} \geqslant 0. \tag{40}$$

Under individual-owned AI, there is an equilibrium (possibly with randomization in time-1 contributions) with full employment in both periods.

The proof is in Section B.7.

The substantive change to the analysis is that, with costly contributions, equilibrium contributions can be strictly higher than in the no-surveillance baseline. This is because workers are rewarded for their marginal contribution to AI quality, and can increase that contribution at a cost.

[17] That is, the following results extend with no modification: Observation 4.1, Theorem 4.2, Observation 4.5, Observation 4.6, Observation 4.7, Theorem 4.8, and Theorem 4.9.

[18] This complication does not arise under firm ownership of data because worker contributions at time-1 are a supermodular game; see Lemma 4.4.

Recall that under firm ownership, equilibrium contributions never exceed θ^j (Observation 4.6). The conclusion that individual ownership raises contributions compared to firm ownership holds even under costly contributions, as we now state formally.

Theorem 4.23. Under any equilibrium with individual ownership, if worker j is hired and contributes k_1^j with positive probability, then $k_1^j \geqslant \theta^j$.

The proof is in Section B.8.

However, the welfare effects of individual ownership can change under costly contributions; Observation 4.12 does not follow without further assumptions. The key is that under individual ownership, workers may make costly contributions that are individually rational but not socially beneficial, as in the following example.

Example 2. There are two workers, with $K^1 = K^2 = \{5, 6\}$. The AI has quality 8 if both workers contribute 6, quality 4 if exactly one worker contributes 6, and quality 0 otherwise. Contribution 5 is free (so $\theta^1 = \theta^2 = 5$), contribution 6 costs $\epsilon = 5$, and $\psi = \gamma = 1$. Under firm ownership, in the unique equilibrium both workers contribute 5; the AI has quality 0 and total surplus equals the no-AI level of 20. Under individual ownership, there is an equilibrium in which both workers contribute 6: doing so raises a worker's time-2 wage from 1 to 7, which exceeds the cost $\epsilon = 5$. In this equilibrium the AI outperforms every worker, so time-2 output rises from 10 to 16, but the contribution costs (10) exceed the combined output gains (2 at time 1 and 6 at time 2): total surplus is 18, strictly lower than under firm ownership, and also strictly lower than in the no-AI case.

Next we study equilibria under the data trust, maintaining the own-gains rule (34). Because each worker's time-2 compensation (34) is non-decreasing in their own contribution, the analysis extends smoothly: under condition (40), there is an equilibrium (possibly with randomization in time-1 contributions) with full employment in both periods—again with no analogue of Condition C, because the argument of Proposition 4.16 applies at every realized contribution profile—and, by the argument of Theorem 4.23, every contribution in the support is at least θ^j. In general, the comparison of total surplus between individual ownership and the data trust is ambiguous. Example 2 proves that individual ownership can sometimes result in strictly lower total surplus than the no-AI case. By contrast, the data trust always raises total surplus compared to the no-AI case, as we now state formally.

Theorem 4.24. Under any equilibrium with the data trust and the own-gains rule (34), expected total surplus is at least as high as in the no-AI case.

The proof is in Section B.9.

5 Conclusion

AI models can now perform a wide range of workplace tasks. But in many cases, model performance depends on access to detailed knowledge about how work is actually done, knowledge that has traditionally belonged to workers. While many foundational AI models were trained on historical records collected without workers' awareness or meaningful consent, further progress will depend on new data from workers who are increasingly aware that their actions, communications, and outputs can be used to train systems that replicate their work.

Our survey of U.S. full-time workers suggests that, even with growing workplace surveillance, workers perceive a substantial knowledge gap between what they know about how to do their jobs well, and what their employers have documented. Training AI systems may require firms to find ways to encourage workers to share their additional expertise.

We provide evidence that workers' have agency in choosing how much knowledge to supply to their employers. This empirical finding motivates our theory, which considers one important reason why workers may want to adjust their knowledge supply: career concerns. We show that, under current institutional arrangements, concerns about expropriation give workers an incentive to hide knowledge. Such behavior can have negative consequences for all parties: workers can suffer lower productivity and wages in the present, firms can face lower profits from reduced ability to make use of labor expertise, and overall economic output can decline due to lower productivity gains from the adoption of effective AI systems. These frictions give workers, employers, and policymakers a shared interest in governance structures that mitigate fears of data-driven expropriation.

Our analysis also reveals a tension between workers' stated preferences and their longer-term welfare. Roughly 65% of respondents favor individual control over their work data, including the right to sell it for AI development. Our theory shows, however, that unilateral sales create a competition externality: a given worker's decision to sell their data improves the firm's outside option against all other workers. Because workers do not internalize this spillover, individual data sales can leave all workers worse off relative to a no-AI benchmark, even when they hold full bargaining power. By contrast, collective data ownership internalizes these externalities and protects work-

ers' compensation for their data, however weak the trust's bargaining position. Implementing such arrangements, however, raises practical challenges. Differences in the value of workers' contributions may generate tensions within the trust over how to divide the proceeds, and the legal basis for worker ownership is difficult to define when data are produced using firm-provided capital and intellectual property.

More broadly, our results suggest that the governance of workplace data will play a central role in shaping the trajectory of AI adoption and worker welfare. If current arrangements persist, rising worker awareness may slow the development of effective AI by inducing meaningful resistance, as evidenced by growing labor disputes around AI in the workplace, or by merely failing to motivate valuable knowledge contributions. Designing, testing, and refining alternative mechanisms, whether through public policies or private contracts, is an important direction for future research.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022). Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics 14*(4), 218–256.

Acquisti, A., C. Taylor, and L. Wagman (2016). The economics of privacy. *Journal of Economic Literature 54*(2), 442–492.

Aitken, H. G. J. (1960). *Taylorism at Watertown Arsenal: Scientific Management in Action, 1908–1915*. Cambridge, MA: Harvard University Press.

Allyn-Feuer, A. and T. Sanders (2023). Transformative AGI by 2043 is <1% likely. *arXiv preprint arXiv:2306.02519*.

Arrieta-Ibarra, I., L. Goff, D. Jiménez-Hernández, J. Lanier, and E. G. Weyl (2018). Should we treat data as labor? moving beyond "free". *AEA Papers and Proceedings 108*, 38–42.

Baker, G., R. Gibbons, and K. J. Murphy (1994). Subjective performance measures in optimal incentive contracts. *Quarterly Journal of Economics 109*(4), 1125–1156.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press.

Bergemann, D. and A. Bonatti (2019). Markets for information: An introduction. *Annual Review of Economics 11*, 85–107.

Bloesch, J., B. Larsen, and B. Taska (2022). Which workers earn more at productive firms? position specific skills and individual worker hold-up power. Working paper.

Boston Consulting Group (2025, September). The widening AI value gap: Build for the future 2025. Technical report, Boston Consulting Group.

Brynjolfsson, E. and Z. Hitzig (2025). AI's use of knowledge in society. Chapter prepared for the NBER volume on the economics of transformative AI, National Bureau of Economic Research.

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

Bubeck, S., V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi, M. T. Ribeiro, and Y. Zhang (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv preprint arXiv:2303.12712*.

Buckman, S. R., J. M. Barrero, N. Bloom, and S. J. Davis (2025, February). Measuring work from home. Working Paper 33508, National Bureau of Economic Research.

Cavallo, A. and Z. Cullen (2026). A market for human expertise: Observations. Mimeo, Harvard Business School.

Chequer, M., K. Herkenhoff, D. Papanikolaou, J. Rothbaum, L. Schmidt, and B. Seegmiller (2026). Labor, despite AI. Working paper.

Christopher, N. (2025, November). Inside the race to train AI robots how to act human in the real world. Los Angeles Times. Available at https://www.latimes.com/business/story/2025-11-02/inside-californias-rush-to-gather-human-data-for-building-humanoid-robots. Accessed 2026-03-18.

Cui, Z. K., M. Demirer, S. Jaffe, L. Musolff, S. Peng, and T. Salz (2025). The effects of generative ai on high skilled work: Evidence from three field experiments with software developers. *Management Science*. Available at SSRN 4945566.

Diamantis, M. E. (2023). Employed Algorithms: A Labor Model of Corporate Liability for AI. *Duke Law Journal 72*, 797–867.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2023). Gpts are gpts: An early look at the labor market impact potential of large language models.

Farboodi, M. and L. Veldkamp (2021). A growth model of the data economy. Working Paper 28427, National Bureau of Economic Research.

Freixas, X., R. Guesnerie, and J. Tirole (1985). Planning under incomplete information and the ratchet effect. *Review of Economic Studies 52*(2), 173–191.

Gibbons, R. (1987). Piece-rate incentive schemes. *Journal of Labor Economics 5*(4), 413–429.

Grace, K., Z. Stein-Perlman, and B. Weinstein-Raun (2022). 2022 Expert Survey on Progress in AI. AI Impacts report (Aug. 2022). Available at https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/.

Grossman, S. J. and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy 94*(4), 691–719.

Grout, P. A. (1984). Investment and wages in the absence of binding contracts: A nash bargaining approach. *Econometrica 52*(2), 449–460.

Hertel-Fernandez, A. (2024, October). Estimating the prevalence of automated management and surveillance technologies at work and their impact on workers' well-being. Technical report, Washington Center for Equitable Growth, Washington, DC.

Hoffman, M. and S. V. Burks (2020). Worker overconfidence: Field evidence and implications for employee turnover and firm profits. *Quantitative Economics 11*(1), 315–348.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *Review of Economic Studies 66*(1), 169–182.

Hu, K. (2025, December). Ai companies need business customers. why that could take a while. *Reuters*.

IBM Institute for Business Value (2025, May). CEO study: Double down on AI while navigating enterprise hurdles. Technical report, IBM.

International Longshoremen's Association (2025). Ila ratifies six-year master contract with nearly 99 https://www.iladistrict.com/ila-ratifies-six-year-master-contract-with-nearly-99-approval-record-wage-increases-automatio

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *American Economic Review 83*(4), 685–709.

Jäger, S. and J. Heining (2022). How substitutable are workers? Evidence from worker deaths. Working Paper 30629, National Bureau of Economic Research.

Jiang, L. Y., X. C. Liu, N. P. Nejatian, M. Nasir-Moin, D. Wang, A. Abidin, K. Eaton, H. A. Riina, I. Laufer, P. Punjabi, M. Miceli, N. C. Kim, C. Orillac, Z. Schnurman, C. Livia, H. Weiss, D. Kurland, S. Neifert, Y. Dastagirzada, D. Kondziolka, A. T. M. Cheung, G. Yang, M. Cao, M. Flores, A. B. Costa, Y. Aphinyanaphongs, K. Cho, and E. K. Oermann (2023). Health system-scale language models are all-purpose prediction engines. *Nature 619*(7969), 357–362.

Jones, C. I. and C. Tonetti (2020). Nonrivalry and the economics of data. *American Economic Review 110*(9), 2819–2858.

Kim, P. T. and R. Leavitt (2026). Data Rights for Workers. *Boston University Law Review*. Forthcoming; Wash. U. Legal Studies Research Paper 25-04-01.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy 117*(5), 914–940.

Massenkoff, M. and P. McCrory (2026). Labor market impacts of AI: A new measure and early evidence.

Milanez, A., A. Lemmens, and C. Ruggiu (2025, February). Algorithmic management in the workplace: New evidence from an oecd employer survey. Technical Report 31, OECD, Paris.

Milgrom, P. and J. Roberts (1990). Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 1255–1277.

Milgrom, P. and C. Shannon (1994). Monotone comparative statics. *Econometrica: Journal of the Econometric Society*, 157–180.

Mitchell, M. (2021). Why AI is harder than we think. *arXiv preprint arXiv:2104.12871*.

Montgomery, D. (1979). *Workers' Control in America: Studies in the History of Work, Technology, and Labor Struggles*. Cambridge: Cambridge University Press.

Murty, R. N. and R. Kumar S (2026, February). When every company can use the same ai models, context becomes a competitive advantage. *Harvard Business Review*.

New Jersey Office of Innovation (2026). Driver's seat cooperative. https://accelerator.fow.nj.gov/participants/drivers-seat-cooperative. Future of Work Accelerator, participant profile. Accessed 13 July 2026.

NVIDIA (2025). GR00T N1: An open foundation model for generalist humanoid robots. Technical report, NVIDIA. arXiv:2503.14734.

Orchi, T. (2023). Driver's seat puts data — and power — in gig workers' hands. https://www.rockefellerfoundation.org/grantee-impact-stories/drivers-seat-puts-data-and-power-in-gig-workers-hands/. The Rockefeller Foundation, Grantee Impact Story, 21 June 2023. Accessed 13 July 2026.

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

Philip Moreira Tomei, B. K. T. (2026, May). What jobs can ai learn? measuring exposure by reinforcement learning. *Arxiv Working Paper*.

Polanyi, M. (1967). *The tacit dimension* (Anchor Books ed. ed.). Doubleday Anchor book, A540. Garden City, N.Y.: Anchor Books.

Posner, E. A. and E. G. Weyl (2018). *Radical Markets: Uprooting Capitalism and Democracy for a Just Society*. Princeton, NJ: Princeton University Press.

Ramani, A. and Z. Wang (2023). Why transformative artificial intelligence is really, really hard to achieve.

SAG-AFTRA (2023). SAG-AFTRA: TV / Theatric Contracts 2023.

SAG-AFTRA (2024). Sag-aftra strikes video games over a.i. https://www.sagaftra.org/sag-aftra-strikes-video-games-over-ai.

SAG-AFTRA (2025, July). SAG-AFTRA members approve 2025 video game agreement. Press release. Accessed: 2026-07-22.

Sankar, P. (2025, November). Just as the data warehouse defined BI, the context layer will define AI. Metadata Weekly (Substack).

Segal, I. (1999). Contracting with externalities. *Quarterly Journal of Economics 114*(2), 337–388.

Shaw, K. and E. P. Lazear (2008). Tenure and output. *Labour Economics 15*(4), 704–723.

Singhal, K., S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, et al. (2023). Large language models encode clinical knowledge. *Nature 620*(7972), 172–180.

Slator (2025). Voice actors to vote on new video game agreement regarding ai speech technology. https://slator.com/voice-actors-to-vote-on-new-video-game-agreement-regarding-ai-speech-technology/.

Staff, M. (2024). At issue in the longshoremen's strike: How much automation is appropriate at ports? Marketplace.

Stole, L. A. and J. Zwiebel (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies 63*(3), 375–410.

Taylor, F. W. (1911). *The Principles of Scientific Management*. New York: Harper & Brothers.

U.S. Congress (1976). Copyright act of 1976, pub. l. no. 94-553, § 101, 90 stat. 2542, 2544. Public Law. Codified at 17 U.S.C. § 101.

U.S. Government Accountability Office (2024, August). Digital surveillance of workers: Tools, uses, and stakeholder perspectives. GAO Report GAO-24-107639, U.S. Government Accountability Office, Washington, D.C.

USMX and ILA (2025). USMX–ILA master contract. https://sanynj.org/wp-content/uploads/2025/04/USMX-ILA-MASTER-CONTRACT-April-4-2025.pdf.

Viljoen, S., D. Li, and Z. Cullen (2026, March). Knowledge guilds: Sharing in the productivity gains from ai. Working paper.

Wall Street Journal (2026, January). Startup hires white-collar workers to train AI to do their old jobs. The Wall Street Journal. Reported via Cybernews, January 14, 2026.

Workers' Algorithm Observatory (n.d.). Driver's seat (sunsetted). https://wao.cs.princeton.edu/. Workers' Algorithm Observatory, Princeton University. Accessed 13 July 2026.

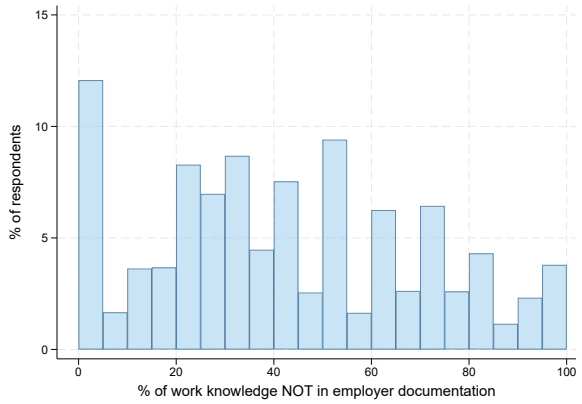
Writers Guild of America West (2023). Artificial intelligence. https://www.wga.org/contracts/know-your-rights/artificial-intelligence. Summary of AI provisions in the 2023 WGA Theatrical and Television Basic Agreement.

Yang, V. and C. Zhou (2026, January). In chinese data factories, workers teach humanoid robots boring tasks. *Rest of World*. Accessed July 2026.

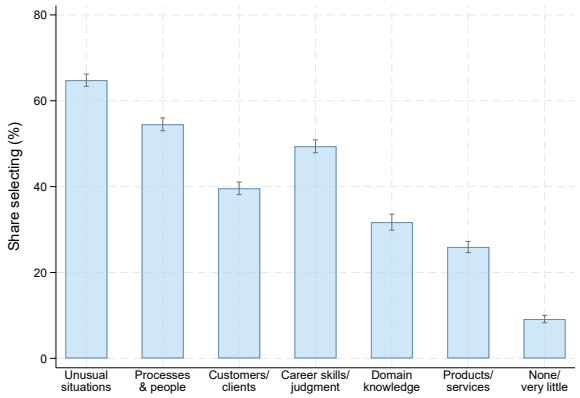
Zhu, Z. and H. Hu (2018). Robot learning from demonstration in robotic assembly: A survey. *Robotics 7*(2), 17.

Figure 1: The Knowledge Gap in the US-Occupations and Partner Firm Surveys

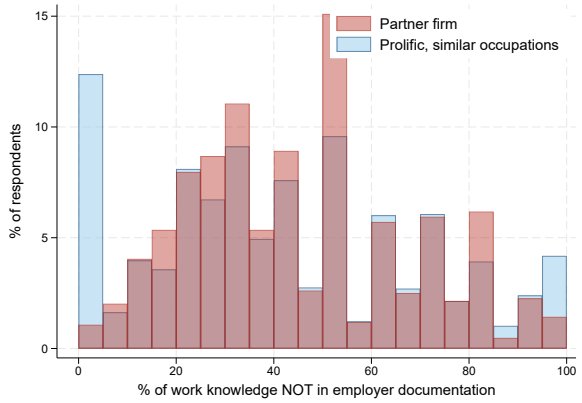
A. Distribution of the Knowledge Gap



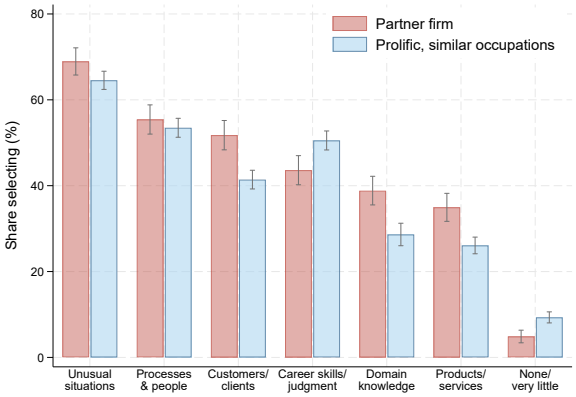
B. Types of Knowledge Not Captured



C. Distribution: Partner-Firm Comparison

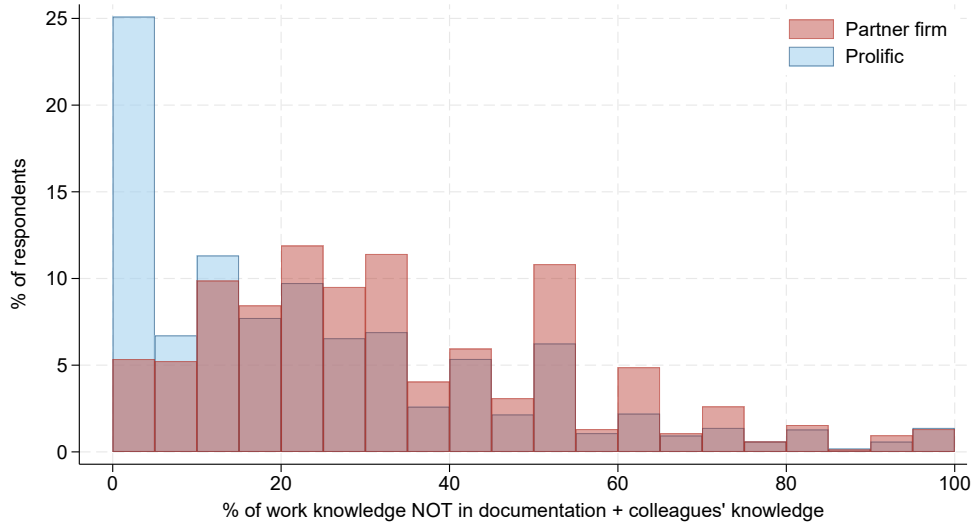


D. Types: Partner-Firm Comparison



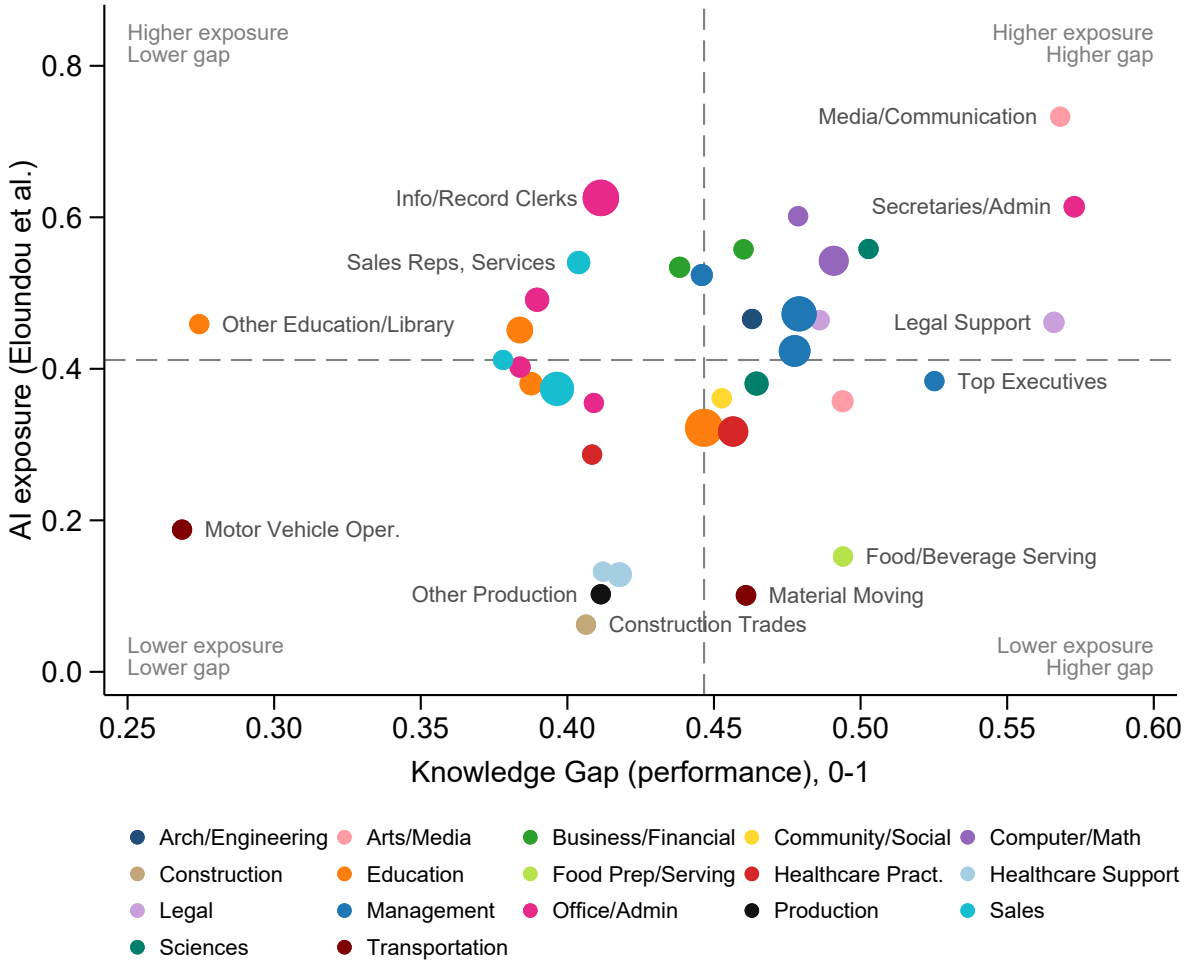
Notes: Panels A and B report the US-Occupations Survey sample (pooled waves, unweighted). Panel A shows the distribution of the Knowledge Gap (Documentation), 100 minus the share of work knowledge the respondent reports is captured by their employer's documentation (N = 4,275); bin width 5. Panel B shows the share of respondents selecting each category in the multi-select question, asked immediately after the documentation question, about which aspects of their work knowledge are not well captured by existing documentation (N = 4,275; the professional domain-knowledge option appears only in the June 30 wave, N = 2,405); the two surveys use near-identical option lists, and category labels are abbreviated; categories are ordered by the partner-firm share in Panel D; whiskers are 95% confidence intervals. Panels C and D compare partner-firm employees (red; all main-survey respondents with a nonmissing response, N = 841, in Panel C; main-sample respondents, N = 821, in Panel D) with the US-Occupations Survey sample (blue) restricted to management, business and financial, sales, and office and administrative occupations, the closest match to the partner firm's workforce composition (N = 1,963; professional domain knowledge, June 30 wave only, N = 1,142). Appendix Figure A2 shows the same occupation-matched comparison for uniquely held knowledge.

Figure 2: The Unique Knowledge Gap in the Partner Firm and US-Occupations Surveys



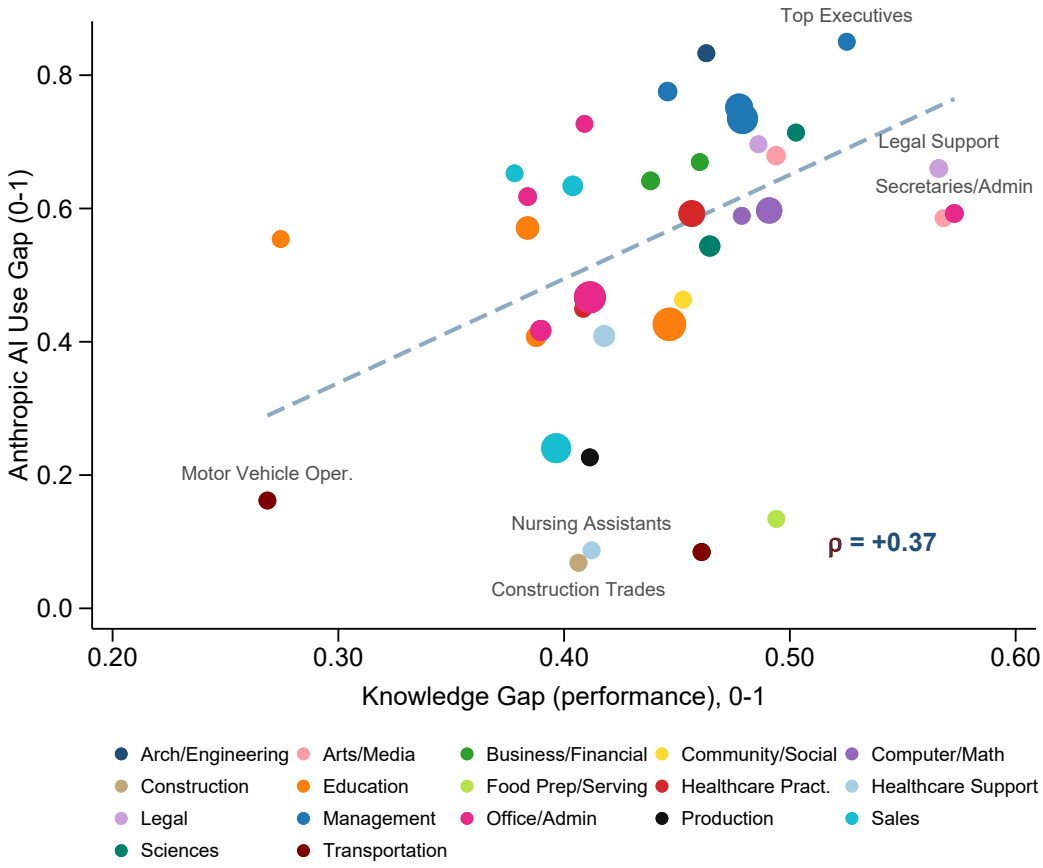
Notes: The figure overlays the distribution of uniquely held knowledge, 100 minus the share of work knowledge the respondent reports is captured by employer documentation plus colleagues' knowledge, for partner-firm employees (red, all main-survey respondents with a nonmissing response, $N = 841$) and US-Occupations Survey respondents (blue, pooled waves, unweighted, $N = 4{,}275$); bin width 5. Partner-firm employees report a mean unique gap of 32.2 (median 29); US-Occupations Survey respondents report a mean of 23.1 (median 19). Appendix Figure A2 repeats the comparison with the US-Occupations Survey sample restricted to occupations matching the partner firm's workforce.

Figure 3: Knowledge Gap vs. AI Exposure



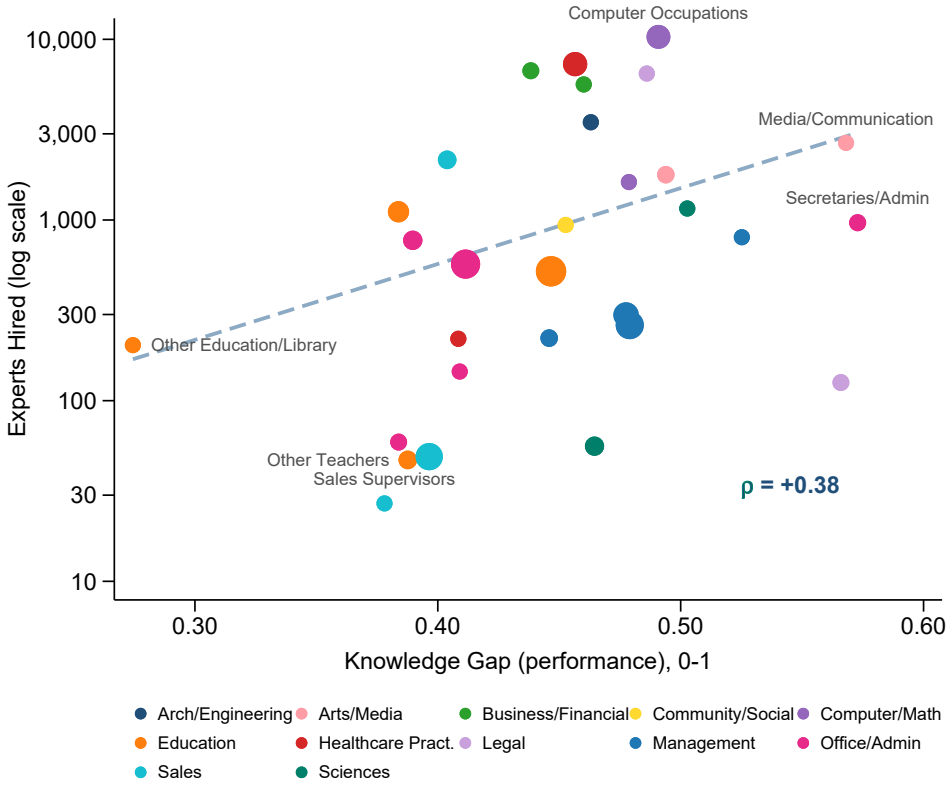
Notes: Each point is a SOC 3-digit minor occupation group with at least 30 survey respondents (37 groups); marker color denotes the parent SOC 2-digit major group and marker size is proportional to the number of survey respondents in the group. The horizontal axis is the occupation's mean replacement gap, the Knowledge Gap (Performance): the self-reported shortfall in a similar replacement's performance when trained only on employer documentation, i.e., 100 minus the response to "*Your replacement is able to learn ONLY from your employer's documentation. On a scale of 0 to 100, how well would your replacement be able to do your job?*", averaged within occupation and rescaled to 0–1. The vertical axis is the occupation's theoretical AI exposure developed by Eloundou et al. (2023), measuring the share of the occupation's work tasks exposed to LLMs; we use the human-rated series, weight tasks by O*NET core importance, and aggregate SOC 6-digit exposure to the SOC 3-digit level using 2025 OEWS employment weights. Both axes are in levels; dashed lines mark the unweighted medians of each measure across the 37 occupation groups, and the corner labels describe the resulting quadrants. Only select groups at the extremities are labeled to limit overplotting.

Figure 4: Knowledge Gap and "Missing" AI Adoption



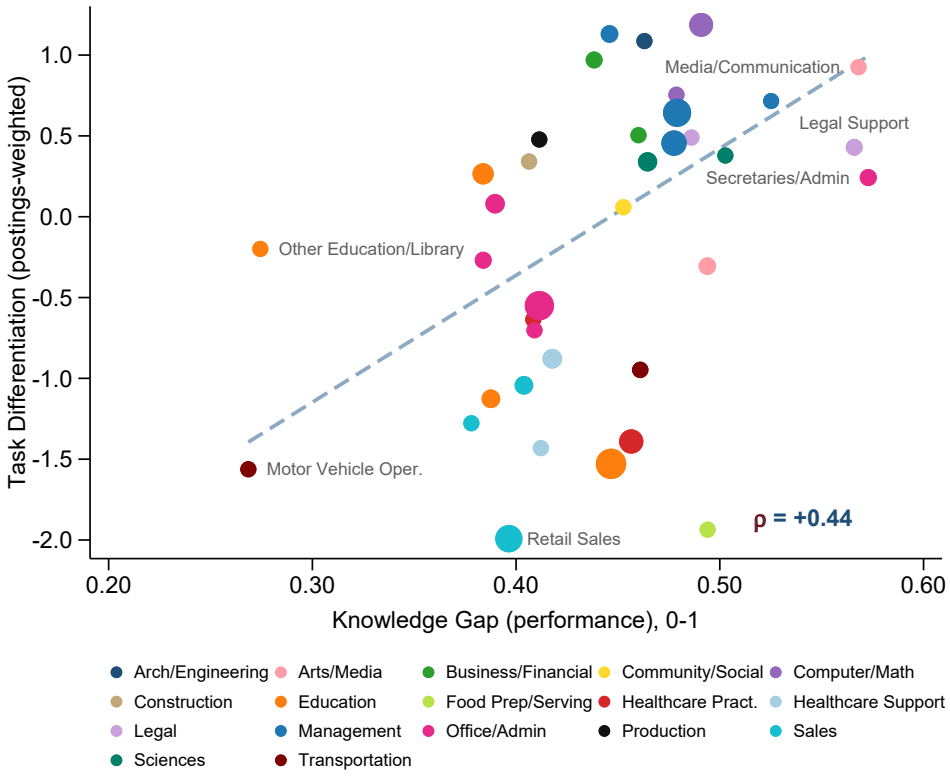
Notes: Each point is one of the 37 SOC 3-digit minor occupation groups; marker color denotes the parent SOC 2-digit major group, marker size is proportional to the number of survey respondents in the group, the dashed line is the OLS fit (weighted by survey respondents), and ρ reports the unweighted correlation. Survey gaps are computed at the individual level and collapsed to the SOC 3-digit occupation mean; only occupation groups with at least 30 respondents are retained. Only select groups at the extremities are labeled to limit overplotting. The figure shows the Anthropic "Missing AI" Index scattered against the occupation's mean replacement gap, defined in Figure 3 and rescaled to 0–1. The "Missing AI" Index, taken from Massenkoff and McCrory (2026), is theoretical AI use minus observed AI use, so higher values mean that more of the conceptually feasible AI use in an occupation has not yet materialized in practice. Correlations are suggestive given the small number of occupation groups.

Figure 5: Knowledge Gap and the Market for Human Expertise



Notes: The figure plots the number of workers hired through Mercor (log scale) against the occupation group's mean replacement gap, defined in Figure 3. Mercor is a leading marketplace through which AI developers hire domain experts to supply training data. Each point is one of the 30 SOC 3-digit minor occupation groups with both Mercor hiring and at least 30 survey respondents; marker color denotes the parent SOC 2-digit major group, marker size is proportional to the number of survey respondents in the group, the dashed line is the OLS fit (weighted by survey respondents), and ρ reports the unweighted correlation, computed on $\log_{10}$ hires. Only select groups at the extremities are labeled to limit overplotting. Hires are measured from daily observations of Mercor's public job board: each posting is classified from its title to a detailed SOC occupation. Because the posted "N hired this month" counter reflects a trailing 30-day window rather than a calendar-month count, a posting's hires are its counter level when first observed plus all subsequent day-over-day increases; occupation totals sum over postings. Appendix Figure A8 presents the Knowledge Gap's relationship to timing, posted prices, and spend.

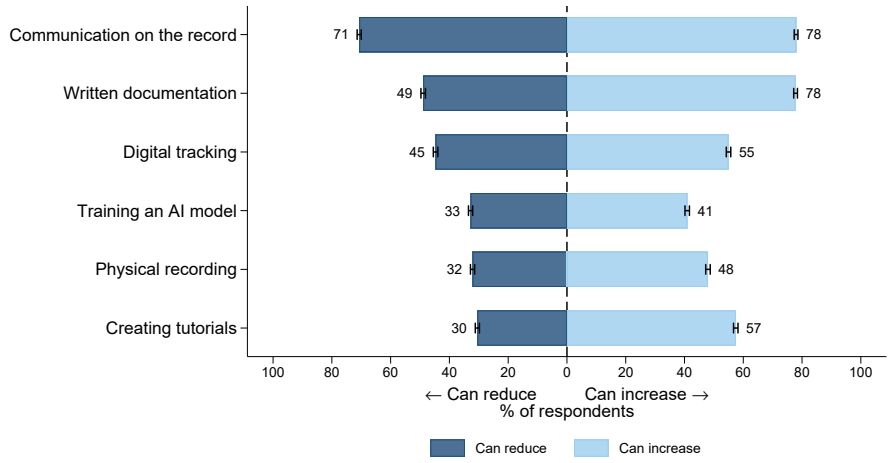
Figure 6: Knowledge Gap and Worker Hold-Up Power



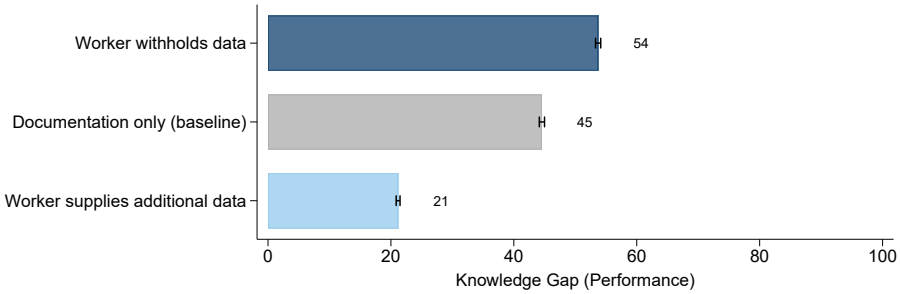
Notes: Each point is a SOC 3-digit minor occupation group with at least 30 survey respondents (37 groups); marker color denotes the parent SOC 2-digit major group, marker size is proportional to the number of survey respondents in the group, the dashed line is the OLS fit (weighted by survey respondents), and ρ reports the unweighted correlation. Only select groups at the extremities are labeled to limit overplotting. The horizontal axis is the occupation's mean replacement gap, defined in Figure 3 and rescaled to 0–1. The vertical axis is the task-differentiation score of Bloesch et al. (2022), a measure of workers' individual hold-up power based on the dissimilarity between a worker's tasks and those of coworkers at the same firm; the score is constructed from job postings at the detailed-occupation level and aggregated to SOC 3-digit groups weighting by posting counts.

Figure 7: Worker Agency over Knowledge Supply

A. Actions Workers Can Take to Adjust Knowledge Supply

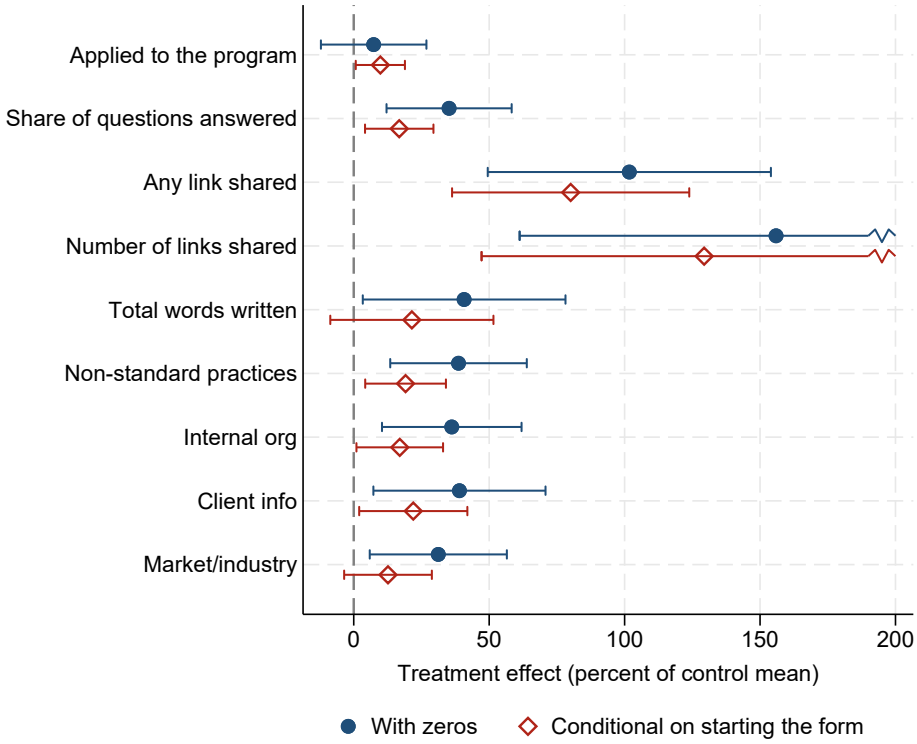


B. Effects of Withholding and Supplying on the Knowledge Gap



Notes: In Panel A, respondents answered two mirrored multi-select questions asking which actions they could take to "*REDUCE*" or "*INCREASE*" "*the amount of data your employer is able to collect about your work activities*" (the increase version adds "*or your thought process behind those activities*"). Dark blue bars give the share selecting an action in the reduce question; light blue bars, the corresponding action in the increase question. Rows: communication on the record; written documentation; digital tracking; physical recording; training an AI model; and creating tutorials. Rows are sorted by the share who could reduce data collection. In Panel B, bars show the mean Knowledge Gap (100 minus the respondent's rating of how well a similar replacement could perform their job) under three data scenarios. The grey bar is the baseline, in which the replacement learns only from the employer's existing documentation. In the withholding scenario (dark blue), the worker individually does everything they can to reduce the amount of data or documentation their employer is able to collect, while coworkers do not change their behavior. In the additional-supply scenario (light blue), the worker individually does everything they can to help the employer create additional data and documentation of their work activities and the thought process behind them, and the replacement learns from both the existing and the new documentation. Whiskers span ± 1 standard error.

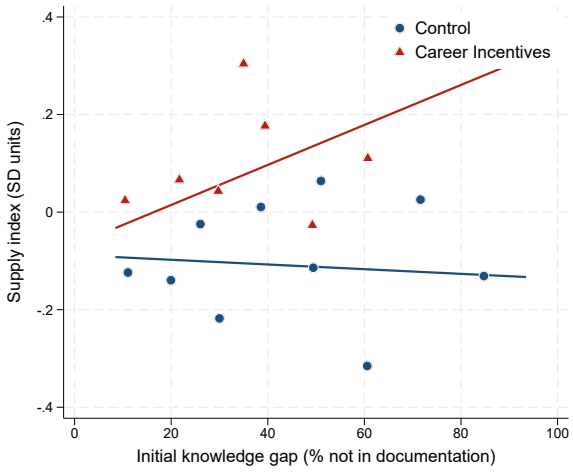
Figure 8: Treatment Effects on Applying and on Knowledge Supply, by Outcome and Content Area



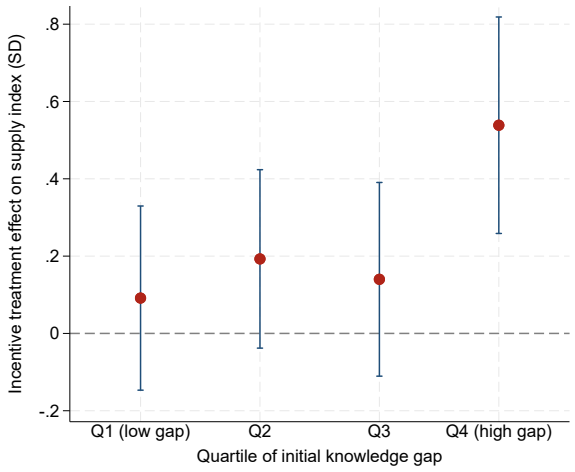
Notes: Incentive treatment effects expressed as a percent of the control-arm mean; whiskers are 95% confidence intervals. Navy circles code non-participation as zero supply; red diamonds condition on reaching the question. For the eight supply outcomes, navy is the main sample (N = 821) with knowledge-assessment non-respondents coded as supplying zero and red is conditional on starting the knowledge assessment (N = 308); both control for pre-randomization intent and AI familiarity fixed effects. For applying (top row), navy is the full randomized frame (N = 5,261: every employee eligible for the experiment, with those who never responded to the survey or did not answer the application question coded as not applying; control mean 10.2%), adjusted for standardized AI-usage metrics (conversations, active days, breadth) and an indicator for having them. Red is conditional on reaching the application question (N = 821, control mean 63.8%). The frame-level effect on applying is +0.8pp (SE 1.0), and the conditional effect is +6.2pp (SE 2.9, p = .035). Content-area outcomes (bottom four) are the share answered of the free-text question family built from the knowledge-capture form: non-standard practices (task-level deviation and hardest-to-explain questions, tool workarounds, hard-won context), internal organization (how decisions and approvals work; making connections, counted only for respondents who do not work mainly independently), client info (block shown only to client-facing respondents: N = 762 with zeros, 249 conditional), and market/industry knowledge. SEs clustered on the incentive randomization unit throughout.

Figure 9: Incentive Effect on Knowledge Supply, by Initial Knowledge Gap and Unique Knowledge

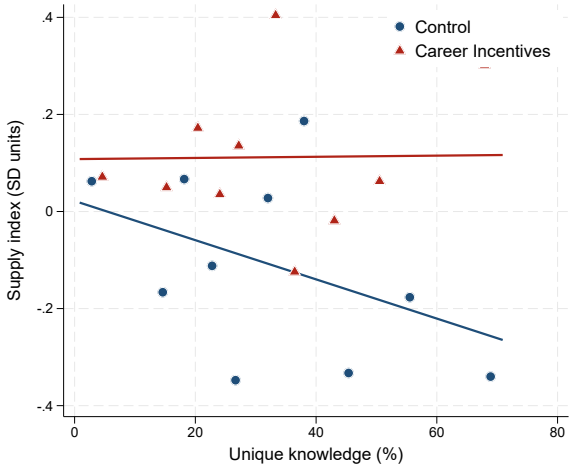
A. Supply Index, by Initial Knowledge Gap



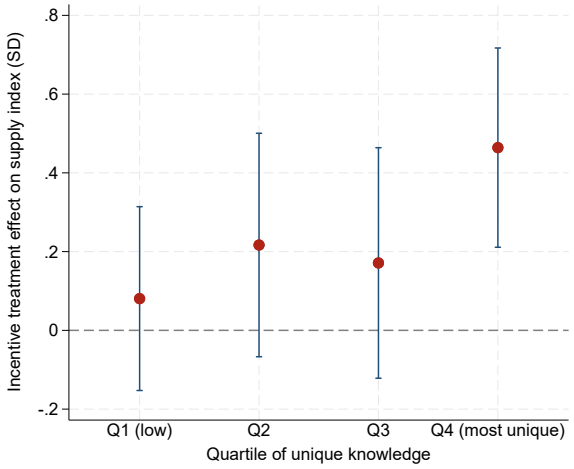
B. Effect, by Quartile of Initial Knowledge Gap



C. Supply Index, by Unique Knowledge



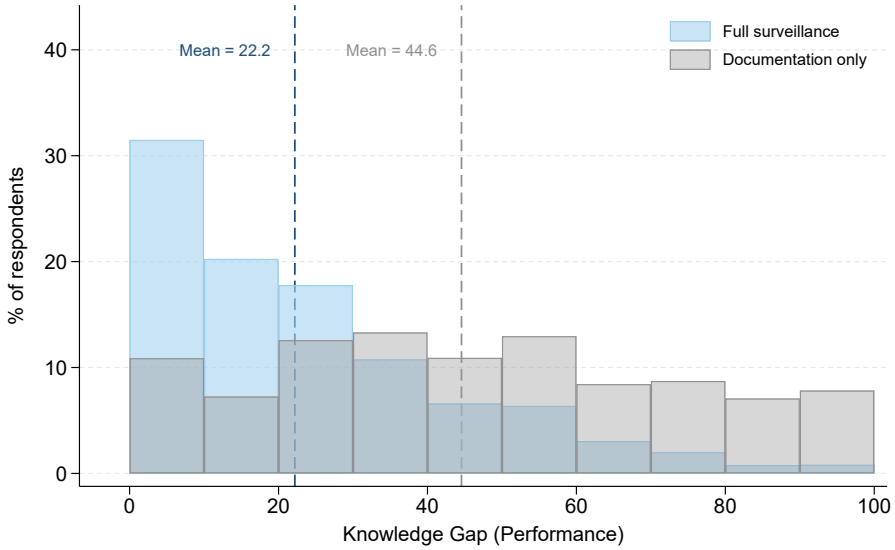
D. Effect, by Quartile of Unique Knowledge



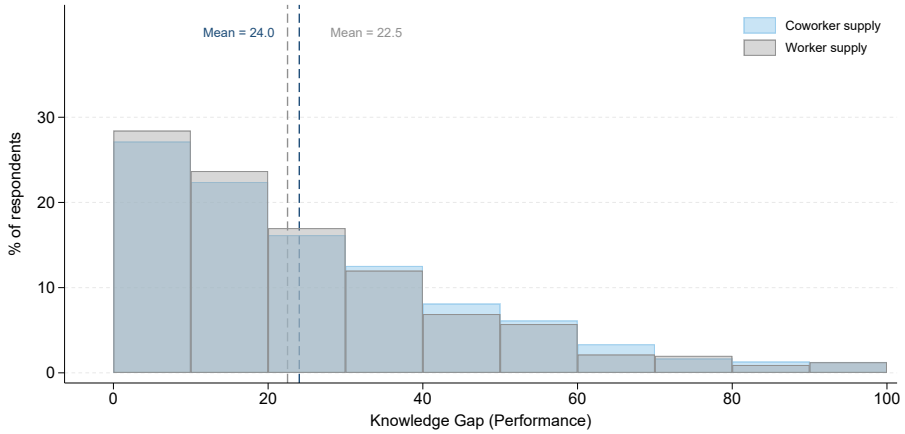
Notes: Partner-firm main sample, both arms ($N = 821$). Panel A: supply index vs. initial knowledge gap, 10 quantile bins per arm; lines are per-arm linear fits; the knowledge gap × incentive interaction is 0.0067 per gap point (SE 0.0028, $p = .016$). Panel B: quartile-specific incentive effects. Panels C and D repeat Panels A and B with unique knowledge in place of the initial knowledge gap, controlling for deciles of the total knowledge gap throughout: in Panel C both axes are residualized on total-gap deciles (grand means added back, so units are natural), and the unique × incentive interaction from the pooled regression with total-gap-decile fixed effects is 0.0049 (SE 0.0031, $p = .11$); in Panel D the quartile-specific effects control for total-gap deciles. All regressions control for pre-randomization intent and AI familiarity fixed effects; SEs clustered on the incentive randomization unit.

Figure 10: Shrinking the Knowledge Gap via Surveillance and Colleague Knowledge Supply

A. Distribution of the Knowledge Gap under Full Surveillance



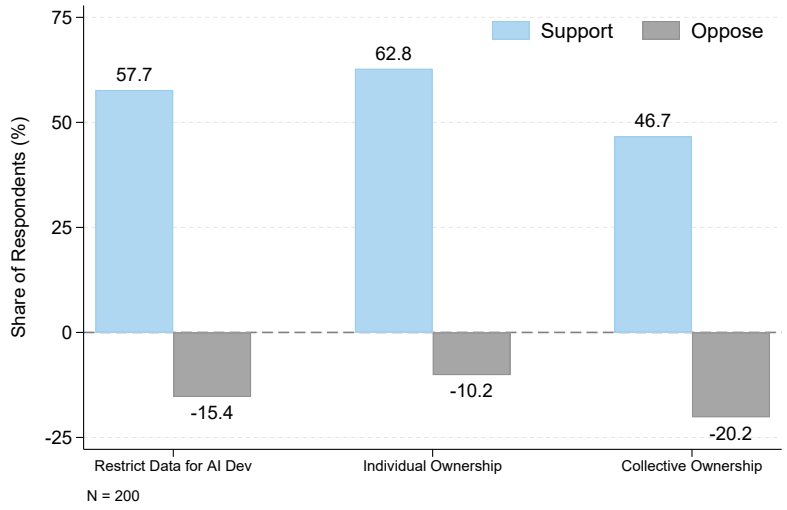
B. Distribution of the Knowledge Gap with Full Colleague Supply



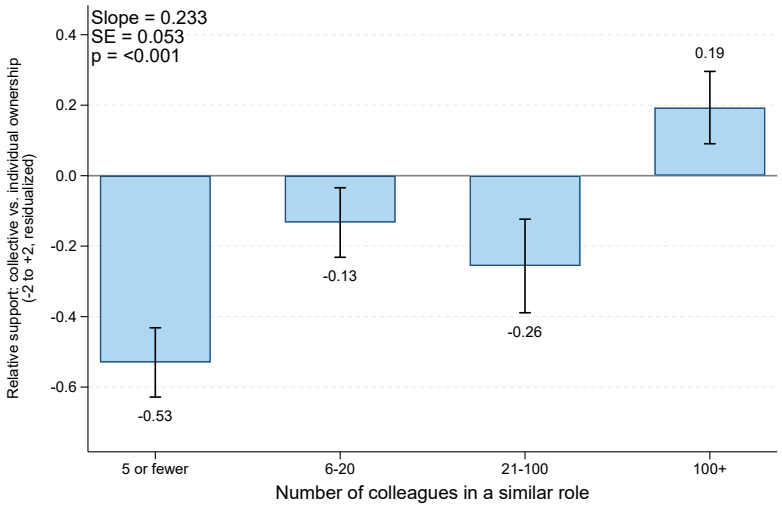
Notes: Each panel overlays two distributions of the Knowledge Gap (100 minus the respondent's rating of how well a similar replacement could perform their job); dashed lines mark the two means. In Panel A, the grey distribution is the baseline, in which the replacement learns only from the employer's existing documentation; the blue distribution is the full-surveillance scenario, in which the replacement also learns from monitoring data covering everything the worker and their coworkers write, say, or do at work (excluding inner thoughts and decision-making). In Panel B, the grey distribution is the worker-supply scenario, in which the replacement learns from the employer's existing documentation plus all additional data and documentation the worker could individually supply; in the blue distribution, it learns from the existing documentation plus all additional data and documentation the worker's colleagues could collectively supply, with the worker's own behavior unchanged.

Figure 11: Policy Preferences

A. Support and Opposition, by Policy



B. Relative Support for Collective vs. Individual Ownership



Notes: In Panel A, bars show the share of respondents who support (blue, upward) or oppose (grey, downward) each of three policies governing work data, presented to respondents as experimental vignettes: a ban on using passively collected work data for AI development; individual ownership of work data, including the right to sell it; and collective ownership, under which workers jointly decide whether to sell their pooled data and how to divide the proceeds. For each policy, respondents were asked "*How would you feel if this policy were in place where you work?*"; support is the share answering "*Positive, I would like it*" and oppose is the share answering "*Negative, I would dislike it.*" "*Neutral*" responses are not shown. The policy questions were administered to a random subsample of respondents ($N = 200$). Section 3.6 describes the policy vignettes in further detail. Panel B plots relative support for collective versus individual ownership (each policy's rating coded $+1/0/-1$ for Positive/Neutral/Negative, then differenced, so the outcome runs from -2 to $+2$) by the number of colleagues the respondent reports working in a similar role, with the four respondents reporting no similar-role colleagues grouped with "5 or fewer." Bars show means of the outcome residualized on SOC 2-digit occupation fixed effects (with the sample mean added back); whiskers are ± 1 standard error of the bin mean. The overlaid slope, standard error, and p-value come from a linear regression on the four-category cohort-size index with the same fixed effects, with standard errors clustered by SOC 2-digit occupation group ($N = 198$).

Table 1: Career-Incentive Effect on Applying

	(1) Applied	(2) Applied
Career Incentives	0.060*	0.062**
	(0.033)	(0.029)
Intent + AI familiarity FE	No	Yes
Control mean	0.635	0.635
Observations	821	821

Notes: Partner-firm main sample, both arms. Linear probability model of applying to the AI Coworker program on the career-incentive treatment; column (2) adds fixed effects for pre-randomization intent to create an AI coworker (1–5) and AI familiarity (1–5). SEs clustered on the incentive randomization unit in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2: Incentive Treatment Effects on Knowledge Supply

	(1) Tasks	(2) Share answered	(3) Log(words)	(4) Links	(5) Supply index
Panel A: Conditional on applying					
Career Incentives	0.310	0.081**	0.555**	0.538***	0.254***
	(0.188)	(0.034)	(0.263)	(0.173)	(0.089)
Intent + AI familiarity FE	Yes	Yes	Yes	Yes	Yes
Control mean	2.004	0.343	2.886	0.381	0.000
Observations	546	546	546	546	546
Panel B: Full main sample					
Career Incentives	0.329**	0.077***	0.547***	0.378***	0.253***
	(0.139)	(0.026)	(0.198)	(0.117)	(0.074)
Intent + AI familiarity FE	Yes	Yes	Yes	Yes	Yes
Control mean	1.273	0.218	1.834	0.242	0.000
Observations	821	821	821	821	821

Notes: Partner-firm main sample. Knowledge-assessment non-respondents are coded as supplying zero in both panels (no non-applier did the assessment, so in Panel B this equals zeroing non-appliers). The supply index is the equal-weighted average of the four standardized components, standardized on control-arm appliers in Panel A and on the full control arm in Panel B (control mean 0 by construction). All columns control for pre-randomization intent (1–5) and AI familiarity (1–5) fixed effects; SEs clustered on the incentive randomization unit in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Incentive Treatment Effects on Knowledge Supply, by Type of Knowledge

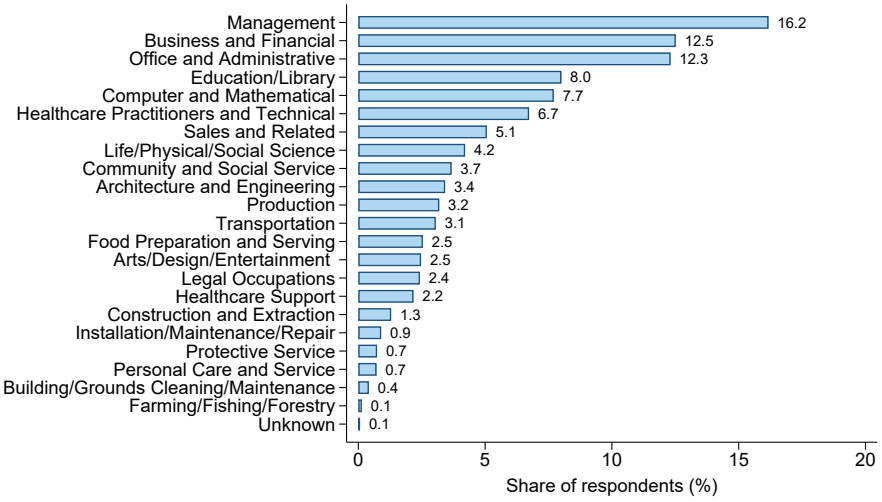
	(1) Non-standard	(2) Internal org	(3) Client info	(4) Market/industry
Panel A: Conditional on applying				
Career Incentives	0.089**	0.083**	0.081*	0.065*
	(0.037)	(0.040)	(0.043)	(0.039)
Intent + AI familiarity FE	Yes	Yes	Yes	Yes
Control mean	0.338	0.351	0.301	0.340
Observations	546	546	487	546
Panel B: Full main sample				
Career Incentives	0.083***	0.081***	0.072**	0.067**
	(0.028)	(0.029)	(0.030)	(0.028)
Intent + AI familiarity FE	Yes	Yes	Yes	Yes
Control mean	0.215	0.223	0.185	0.216
Observations	821	821	762	821

Notes: Partner-firm main sample. Outcomes are the share answered of each free-text question family from the knowledge-capture form: non-standard practices (task-level deviation and hardest-to-explain questions, tool workarounds, hard-won context), internal organization (decisions and approvals; making connections, counted only for respondents who do not work mainly independently), client info (shown only to client-facing respondents; respondents in internal roles are missing, so column 3 has fewer observations), and market/industry knowledge. Knowledge-assessment non-respondents are coded as zero in both panels. All columns control for pre-randomization intent (1–5) and AI familiarity (1–5) fixed effects; SEs clustered on the incentive randomization unit in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

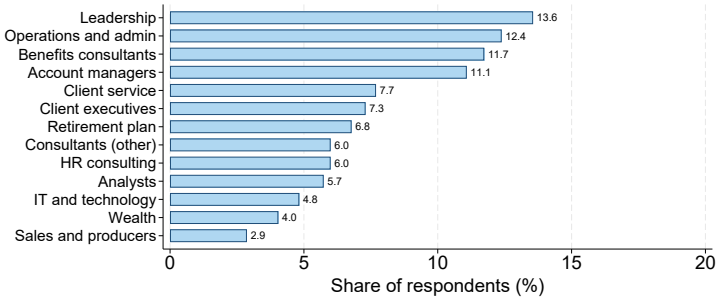
A Appendix Tables and Figures

Figure A1: Distribution of Respondents across Occupations

A. US-Occupations Survey Sample

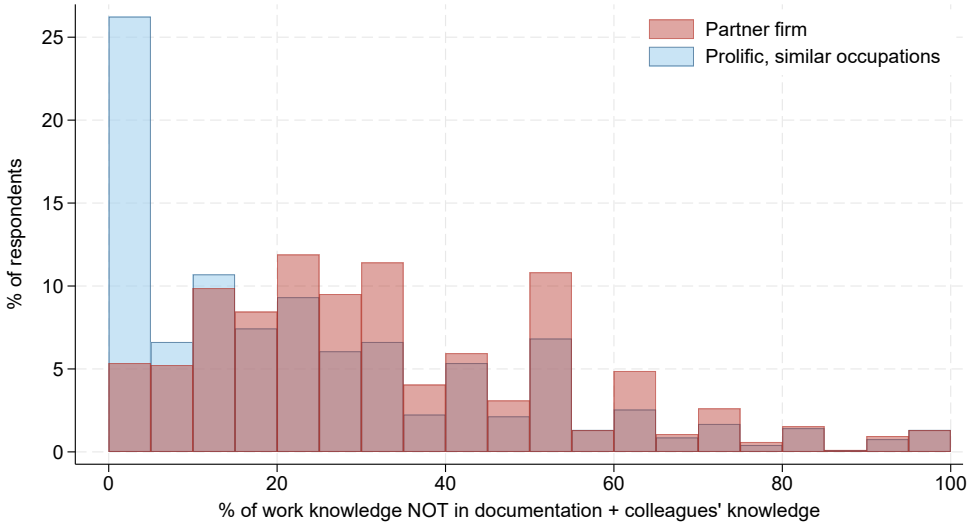


B. Partner Firm Sample



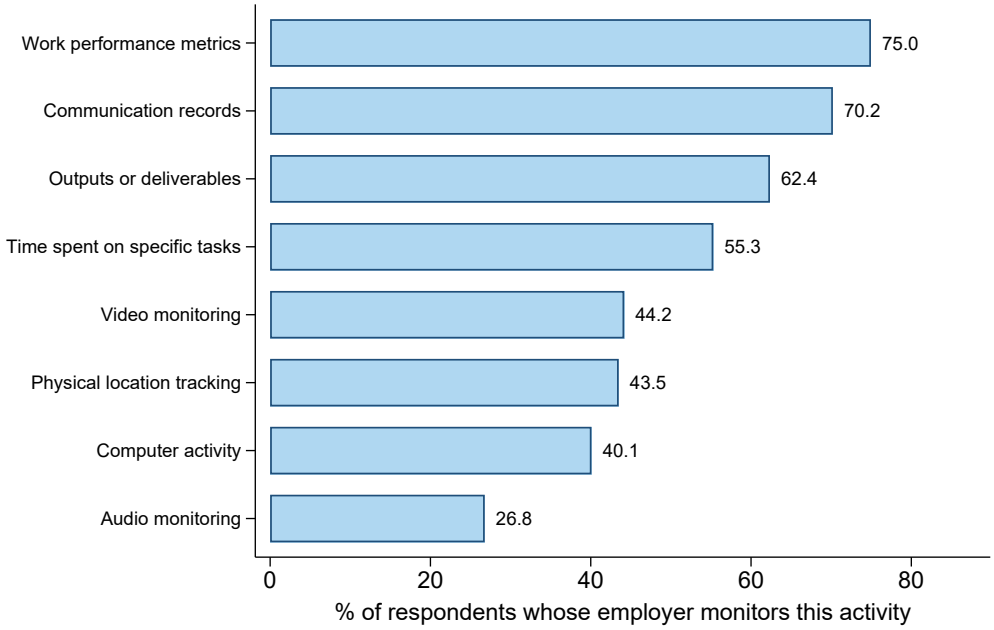
Notes: Panel A shows the share of the US-Occupations Survey main analysis sample in each SOC 2-digit occupation group, sorted from largest to smallest. The sample comprises currently full-time employed U.S. workers recruited through Prolific and surveyed about their primary job. Recruitment was stratified across economic sectors, with sector-level quotas set so that the sample spans the full occupational distribution of U.S. full-time work rather than Prolific's default, which over-represents computer-based occupations. The figure reports unweighted shares; within strata, we construct post-stratification weights that match the weighted sample to the Census distribution of full-time workers on demographics (sex, age, and education) within major (SOC 2-digit) occupation groups, and unless noted otherwise, figures and tables report weighted statistics. The most common groups are management (15.9%), business and financial operations (13.0%), and office and administrative support (12.1%). Panel B shows the corresponding distribution for the partner-firm sample across functional title-team groups, assigned from job titles; groups with fewer than 20 respondents are omitted, and shares are of respondents in the displayed groups ($N = 767$). The most common groups are leadership (13.6%), operations and administration (12.4%), and benefits consultants (11.7%).

Figure A2: Occupation-Matched Comparison of Uniquely Held Knowledge



Notes: The figure repeats Figure 2 with the US-Occupations Survey sample restricted to management, business and financial, sales, and office and administrative occupations (N = 1,963 with nonmissing gaps), the closest match to the partner firm's workforce composition. The corresponding occupation-matched comparison for the documentation gap appears in Panels C and D of Figure 1.

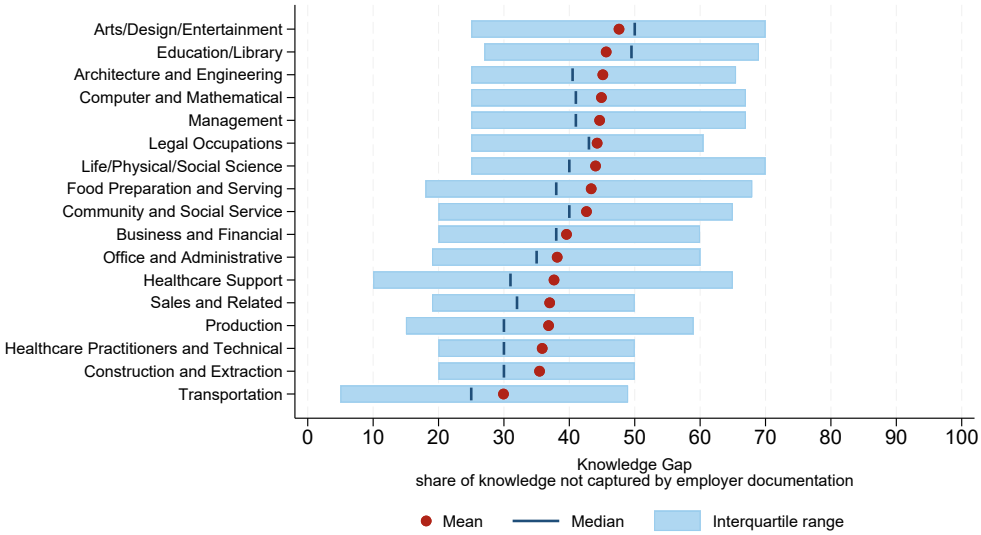
Figure A3: Prevalence of Workplace Monitoring, by Type



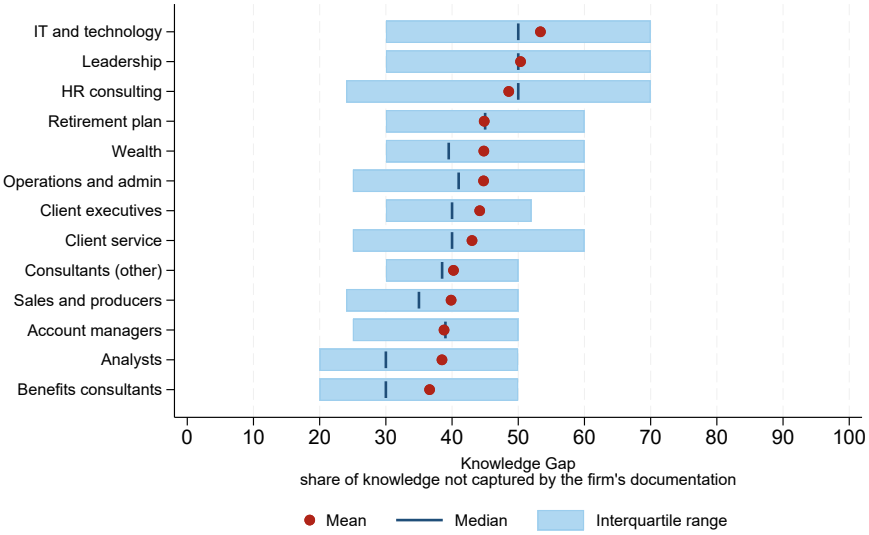
Notes: Share of respondents reporting that their employer monitors or collects data on each activity, from the multi-select question "*Which of the following activities does your employer monitor or collect data on?*" Response options for each activity were Yes / No / Not sure; bars report the share answering Yes, with No and Not sure responses included in the denominator. The count of Yes responses across these eight channels is the surveillance-intensity measure (0–8) used in Appendix Figure A9 .

Figure A4: Distribution of the Knowledge Gap, by Occupation

A. US-Occupations Survey Sample



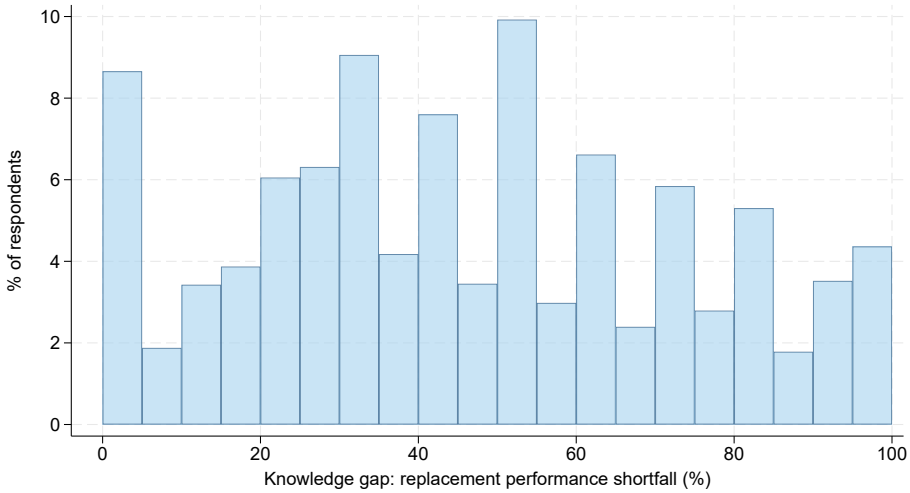
B. Partner Firm Sample



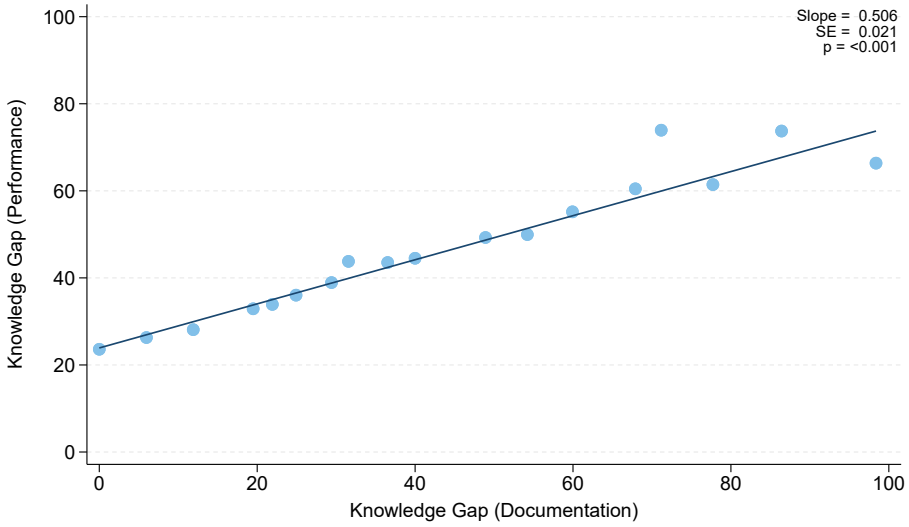
Notes: Both panels show the distribution of the Knowledge Gap (Documentation), 100 minus the share of work knowledge the respondent reports is captured by their employer's documentation, by occupation group sorted by the group mean; red dots mark group means, vertical ticks medians, and shaded bands interquartile ranges. Panel A groups US-Occupations Survey respondents by SOC 2-digit occupation; groups with fewer than 40 respondents are omitted. Panel B groups partner-firm respondents by functional title-team (keyword grouping of job titles); groups with fewer than 20 respondents are omitted.

Figure A5: The Knowledge Gap (Performance) at the Person Level

A. Distribution of the Knowledge Gap (Performance)

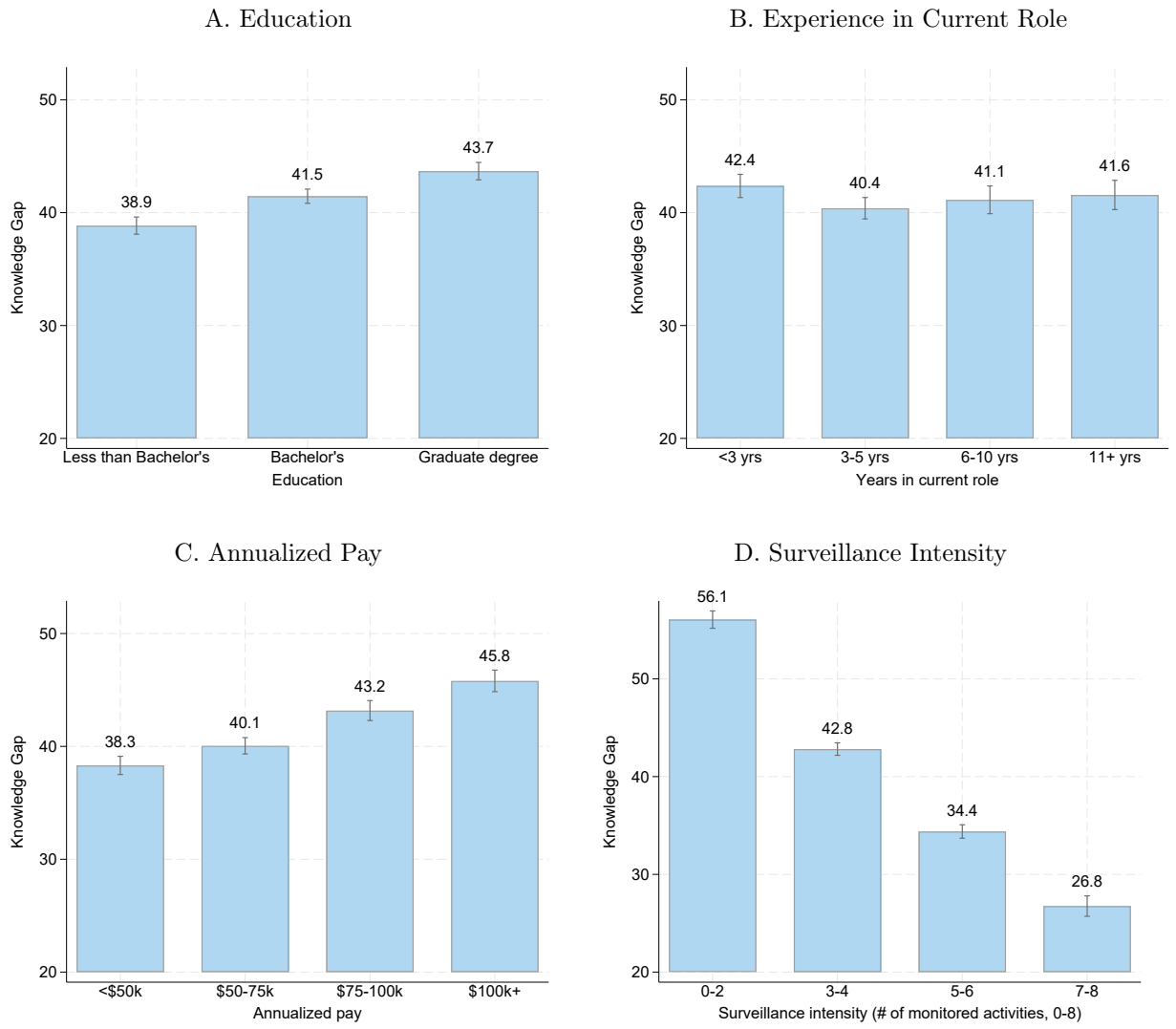


B. Correlation of the Knowledge Gap (Performance) vs. the Knowledge Gap (Documentation)



Notes: Panel A plots the person-level distribution of the Knowledge Gap (Performance): 100 minus the respondent's rating of how well a replacement trained only on the employer's documentation could perform their job (pooled US-Occupations Survey waves, unweighted, N = 4,262; bin width 5). Panel B is a binned scatterplot of the Knowledge Gap (Performance) against the Knowledge Gap (Documentation) at the individual level, both as defined in Appendix Figure A4: the documentation measure is the share of work knowledge not captured by the employer's documentation, and the performance measure is the shortfall in how well a replacement trained only on that documentation could perform the worker's job. Observations are grouped into 20 equal-sized bins of the documentation measure; each point plots the mean of both measures within a bin, and the line is the linear fit estimated on the underlying individual-level data using survey weights. The overlaid slope, standard error, and p-value come from that regression, with robust standard errors.

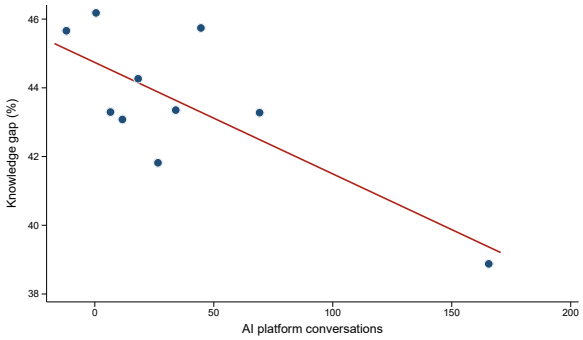
Figure A6: Within-Occupation Correlates of the Knowledge Gap



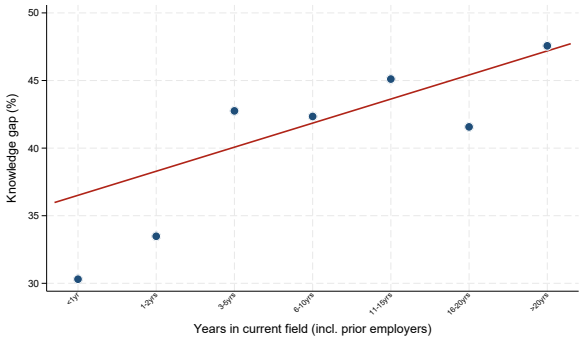
Notes: Each panel plots the mean Knowledge Gap, 100 minus the share of work knowledge the respondent reports is captured by their employer's documentation, by categories of a worker characteristic. The gap is demeaned by SOC 2-digit major group and the grand mean is added back; whiskers show ± 1 standard error of the cell mean. Panel B uses the June 30 wave, the only wave asking years in current role (N = 2,451); the other panels pool both US-Occupations Survey waves. Annualized pay is annual base salary for salaried respondents and hourly wage $\times$ 2,080 hours for hourly respondents. Surveillance intensity is the number of the eight monitoring channels in Appendix Figure A3 that the respondent reports (binned). In Panel D, the gap is additionally residualized on fixed effects for education and deciles of annualized pay (grand mean added back), so the surveillance gradient is net of these worker characteristics.

Figure A7: Partner-Firm Knowledge-Gap Measures: AI Use and Field Experience

A. Knowledge Gap and Observed AI Context

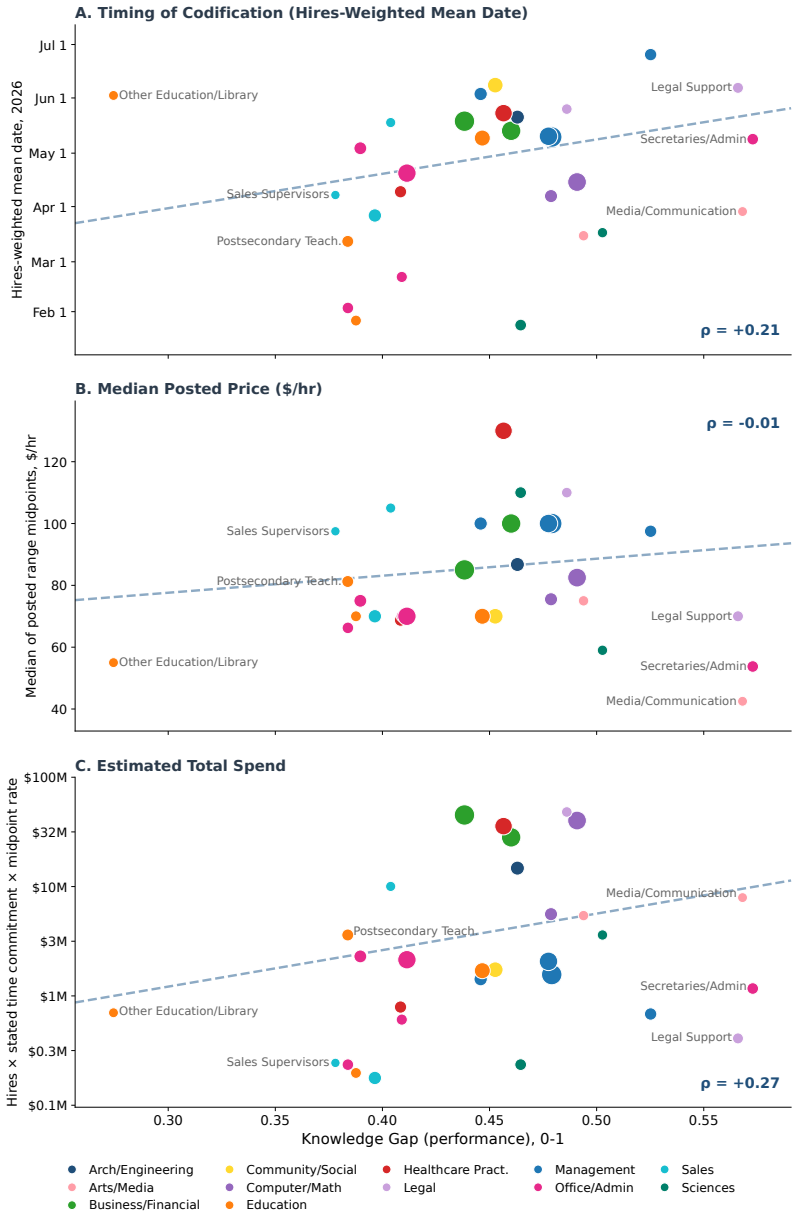


B. Knowledge Gap and Field Experience



Notes: Both panels use partner-firm main-sample respondents. Panel A: respondents with platform AI-usage metrics ($N = 818$; 802 in cells with team variation). Observed AI context is the count of the employee's conversations on the firm's AI platform (administrative, pre-treatment). Both axes are residualized on team fixed effects (primary practice × division) with grand means added back, so units are natural. Dots are decile-bin means; the line is the linear fit: -1.9 gap points per SD of conversations (≈ 60; robust SE 0.66, $p = .003$). Panel B: dots are mean knowledge gaps by years of experience in the respondent's current field, including time at prior employers ($N = 777$ with a nonmissing response); the line is the linear fit across the seven experience categories (+1.8 gap points per category; robust SE 0.46, $p < .001$). The field-experience question is asked in the survey's closing block, after the treatment screens; the pattern is unchanged when restricted to the control arm.

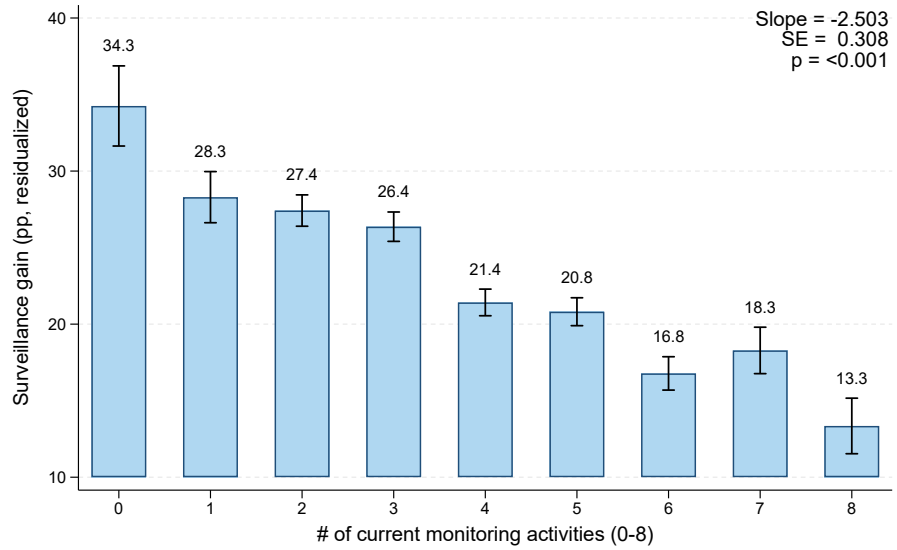
Figure A8: A Market for Human Expertise: Mercor Hiring, Prices, and the Knowledge Gap



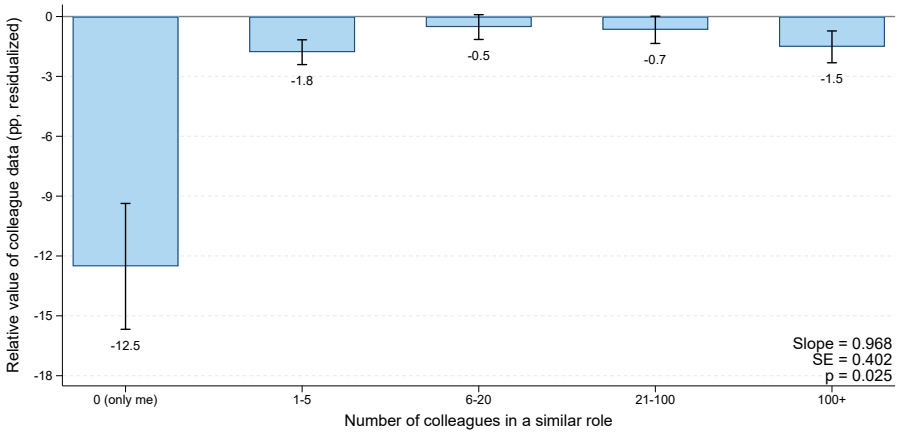
Notes: Data are daily observations of Mercor's public job board; each posting is classified from its title to a detailed SOC occupation and aggregated to the 3-digit minor group. Each point is one of the 30 SOC 3-digit minor occupation groups matched between the survey and Mercor hiring, restricted to groups with at least 30 survey respondents; marker color denotes the parent SOC 2-digit major group, marker size is proportional to survey respondents, dashed lines are OLS fits weighted by survey respondents, and ρ reports the unweighted correlation, computed on $\log_{10}$ of the plotted variable where its axis is log-scaled. The replacement gap on the horizontal axis, the Knowledge Gap (Performance) defined in Figure 3, measures the shortfall in how well a replacement could perform with access only to documentation. Panel A: the timing of codification, measured as the hires-weighted mean calendar date of the group's Mercor hiring. Panel B: the median across the group's hourly postings of the posted-range midpoint; project-priced postings without an hourly rate are excluded. Panel C: estimated total spend = each posting's hires × its stated time commitment × its posted midpoint rate, summed over the group's postings, on a log scale. Sales Supervisors, Other Educational Instruction and Library, Retail Sales, Life/Physical/Social Science Technicians, and Material Recording each have fewer than 100 Mercor hires over the panel.

Figure A9: Heterogeneity in the Returns to Surveillance and Colleague Data

A. Surveillance Gain, by Current Monitoring Intensity



B. Relative Value of Colleague Data, by Workplace Cohort Size



Notes: In both panels, bars show means of the outcome residualized on SOC 2-digit occupation and NAICS 2-digit industry fixed effects (with the sample mean added back); whiskers are ±1 standard error of the bin mean, and the overlaid slope, standard error, and p-value come from the corresponding linear regression with the same fixed effects, with standard errors clustered by SOC 2-digit occupation group. Panel A plots the surveillance gain, the reduction in the Knowledge Gap under full surveillance relative to documentation alone (in percentage points, as in Figure 10, Panel A), against the firm's current monitoring intensity: the number of "*Yes*" responses across the eight data-collection channels in Appendix Figure A3 ("*Not sure*" responses are counted as not monitored); N = 4,002. Panel B plots the relative value of colleague data, the difference in rated replacement performance between the colleague-supply and worker-supply scenarios of Figure 10, Panel B, by the number of colleagues the respondent reports working in a similar role at their employer; the regression uses the five-category cohort-size index; N = 2,142.

Table A1: Comparison of Survey and CPS Sample Characteristics

	Survey	CPS	Difference
Age	37.07	42.15	-5.08
	(10.36)	(13.23)	
Male	0.403	0.546	-0.142
White	0.690	0.750	-0.060
Black	0.133	0.133	0.000
Asian	0.069	0.078	-0.010
Private Sector	0.831	0.834	-0.003
Unionized	0.153	0.110	0.043
Salaried	0.533	0.495	0.038
Hourly	0.467	0.505	-0.038
Annual Salary ($)	89,307.00	106,729.88	-17,422.88
	(45,787.26)	(103,062.49)	
Hourly Wage ($)	26.64	35.46	-8.83
	(25.64)	(19.76)	
N	4,043	50,313	

Notes: This table compares summary statistics between our survey sample and a cross-section of U.S. full-time workers, drawn from the 2025 Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). The CPS sample is restricted to employed, full-time wage and salary workers aged 18 and older. "Private Sector" comprises for-profit and non-profit employers, as opposed to government. Self-reported survey earnings are screened for unit-entry errors: hourly wages above $1,000 and annual salaries below $15,000 are set to missing. Standard deviations for continuous variables are shown in parentheses. CPS demographic statistics are weighted with ASEC person weights; union status, the salaried/hourly classification, hourly wage, and annual salary are computed over the CPS earner (Outgoing Rotation Group) subset, approximately one-quarter of the sample, and are weighted with the ORG earnings weight. CPS annual salary reflects prior-year wage income.

B Proofs

B.1 Proof of Theorem 4.9

Proof. The worker payoffs (7) from time-1 contributions induced by each technology are

$$w_1^j - c^j(k_1^j) + \psi\gamma \left\lfloor \theta^j - \alpha^j \left(k_1^{\overline{J}_1 \cup \{j\}} \right) \right\rfloor_+, \tag{41}$$

$$w_1^j - c^j(k_1^j) + \psi\gamma \left\lfloor \theta^j - \tilde{\alpha}^j \left(k_1^{\overline{J}_1 \cup \{j\}} \right) \right\rfloor_+. \tag{42}$$

By the same argument as in Lemma 4.4, these define a pair of supermodular games. Let us index these games by parameter τ, with $\tau = 0$ corresponding to $\left(\alpha^j\right)_{j \in J}$ and $\tau = 1$ corresponding to $\left(\tilde{\alpha}^j\right)_{j \in J}$. By (25), (26), and Lemma 4.3, it follows that each worker j's utility has increasing differences in (k_1^j, τ). Thus, the comparison of extremal equilibria follows by Theorem 6 of Milgrom and Roberts (1990).

Finally, observe that worker j's time-2 wage under the $\overline{k}_1$ and α is

$$\gamma \left\lfloor \theta^j - \alpha^j \left(\overline{k}_1 \right) \right\rfloor_+ \geqslant \gamma \left\lfloor \theta^j - \alpha^j \left(\overline{\tilde{k}}_1 \right) \right\rfloor_+ \geqslant \gamma \left\lfloor \theta^j - \tilde{\alpha}^j \left(\overline{\tilde{k}}_1 \right) \right\rfloor_+, \tag{43}$$

where the first inequality is by α non-decreasing and the second inequality by (25). The right-hand side of (43) is worker j's time-2 wage under the $\overline{\tilde{k}}_1$ and $\tilde{\alpha}$. The same argument holds comparing wages under $\underline{\tilde{k}}_1$ and $\underline{k}_1$. □

B.2 Proof of Proposition 4.11

Proof. We verify that the following assessment is an equilibrium: full employment in both periods, knowledge contributions $k_1^j = k_2^j = \theta^j$, and time-2 wages $\gamma\lambda_j$ as in (27). As argued in the text, each λ_j is non-negative, so full employment at time 2 is consistent with tie-breaking in favor of hiring, and each worker's contribution $k_1^j = \theta^j$ is a best response. It remains to prove that full employment at time 1 is consistent with Nash-in-Nash bargaining; that is, that the pairwise surplus from employing each worker j at time 1 is non-negative.

Fix a worker j. Contributions equal $k_1 = \theta$, so no worker incurs contribution costs, because $c^l(\theta^l) = 0$. Since Condition C restricts profiles with $k_1^l \geqslant \theta^l$ for all l, it applies to k_1. We abbreviate

$$A = \alpha\left(k_1^J\right), \qquad A^{-l} = \alpha\left(k_1^{J\setminus\{l\}}\right), \qquad A^{-jl} = \alpha\left(k_1^{J\setminus\{j,l\}}\right), \tag{44}$$

and note that $A \geqslant A^{-l} \geqslant A^{-jl}$ and $A \geqslant A^{-j} \geqslant A^{-jl}$, by α monotone non-decreasing.

We first compute the pairwise surplus. Consider the bilateral negotiation between the firm and worker j at time 1, holding the other workers' time-1 wages and contributions fixed. If worker j is employed, the time-1 hired set is J, and the joint payoff of the firm and worker j is

$$\sum_{l\in J}\theta^l - \sum_{l\in J} w_1^l + w_1^j + \psi\left(\sum_{m\in J}\max\left\{\theta^m, A\right\} - \gamma\sum_{l\in J}\lambda_l\left(k_1, J\right) + \gamma\lambda_j\left(k_1, J\right)\right). \tag{45}$$

If worker j is not employed, the time-1 hired set is $J\backslash\{j\}$. All workers are still employed at time 2, but worker j's data is absent from the dataset, because ownership requires employment in both periods. The joint payoff is

$$\sum_{l\neq j}\theta^l - \sum_{l\neq j} w_1^l + \psi\left(\sum_{m\in J}\max\left\{\theta^m, A^{-j}\right\} - \gamma\sum_{l\in J}\lambda_l\left(k_1, J\backslash\{j\}\right) + \gamma\lambda_j\left(k_1, J\backslash\{j\}\right)\right). \tag{46}$$

Worker j's time-1 wage w_1^j cancels within the pair, the other workers' time-1 outputs and wages are common to both branches, and worker j's time-2 wage appears once as a cost to the firm and once as income to worker j, so it cancels within each branch. Subtracting (46) from (45), the pairwise surplus is

$$S_j = \theta^j + \psi\left(\sum_{m\in J}\left(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\}\right) - \gamma\sum_{l\neq j}\left(\lambda_l\left(k_1, J\right) - \lambda_l\left(k_1, J\backslash\{j\}\right)\right)\right). \tag{47}$$

Observe that every dataset appearing in (47)—namely J, $J\backslash\{j\}$, $J\backslash\{l\}$, and $J\backslash\{j,l\}$—is missing at most two workers: one removal from the time-1 disagreement with worker j, and at most one more from the time-2 disagreement embedded in another worker's wage.

Next, partition the workers into

$$U = \left\{m \in J : \theta^m \leqslant \alpha\left(k_1^{J\backslash\{l,l'\}}\right) \text{ for all } l, l' \in J\right\}, \qquad L = \left\{m \in J : \theta^m \geqslant A\right\}. \tag{48}$$

Condition C states that $U \cup L = J$; workers satisfying both conditions may be assigned to either cell. Let D be any dataset missing at most two workers, say $J\backslash\{l,l'\} \subseteq D$, allowing $l = l'$. For $m \in U$, we have $\theta^m \leqslant \alpha(k_1^{J\backslash\{l,l'\}}) \leqslant \alpha(k_1^D)$, so $\max\{\theta^m, \alpha(k_1^D)\} = \alpha(k_1^D)$. For $m \in L$, we have $\alpha(k_1^D) \leqslant A \leqslant \theta^m$, so $\max\{\theta^m, \alpha(k_1^D)\} = \theta^m$. Thus, on every dataset appearing in (47), each position's maximum resolves the same way.

Consider the first sum in (47). Each $m \in L$ contributes $\theta^m - \theta^m = 0$, and each $m \in U$ contributes $A - A^{-j}$, so

$$\sum_{m \in J} \left(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\} \right) = |U| \left(A - A^{-j} \right) \geqslant 0. \tag{49}$$

Consider the second sum. Fix $l \neq j$ and expand (27) term by term. The cross terms corresponding to positions $m \in L$ vanish, and each cross term corresponding to a position $m \in U \backslash \{l\}$ equals $A - A^{-l}$, so

$$\lambda_l\left(k_1, J\right) = \max\left\{\theta^l, A\right\} - A^{-l} + |U \backslash \{l\}| \left(A - A^{-l} \right). \tag{50}$$

Likewise, with time-1 hired set $J \backslash \{j\}$, the dataset is $J \backslash \{j\}$ upon agreement with worker l at time 2, and $J \backslash \{j, l\}$ upon disagreement. The cross terms run over positions $m \in J \backslash \{l\}$, *including* position j: worker j is employed at time 2 in this branch, and position j's term is governed by worker j's cell in the partition (48), like any other position's. Hence

$$\lambda_l\left(k_1, J \backslash \{j\}\right) = \max\left\{\theta^l, A^{-j}\right\} - A^{-jl} + |U \backslash \{l\}| \left(A^{-j} - A^{-jl} \right), \tag{51}$$

with the same count $|U \backslash \{l\}|$ as in (50). If $l \in U$, then $\max\{\theta^l, A\} = A$, $\max\{\theta^l, A^{-j}\} = A^{-j}$, and $|U \backslash \{l\}| = |U| - 1$, so

$$\lambda_l\left(k_1, J\right) - \lambda_l\left(k_1, J \backslash \{j\}\right) = |U| \left(\left(A - A^{-l} \right) - \left(A^{-j} - A^{-jl} \right) \right). \tag{52}$$

If $l \in L$, then $\max\{\theta^l, A\} = \max\{\theta^l, A^{-j}\} = \theta^l$ and $|U \backslash \{l\}| = |U|$, so

$$\lambda_l\left(k_1, J\right) - \lambda_l\left(k_1, J \backslash \{j\}\right) = \left(A^{-jl} - A^{-l} \right) + |U| \left(\left(A - A^{-l} \right) - \left(A^{-j} - A^{-jl} \right) \right). \tag{53}$$

Applying the substitutes assumption (2) to worker l's contribution—raising k_1^l from 0 to θ^l, at the pair of base profiles $k_1^{J \backslash \{j,l\}} \leqslant k_1^{J \backslash \{l\}}$—yields

$$A^{-j} - A^{-jl} \geqslant A - A^{-l}; \tag{54}$$

that is, worker l's data has a weakly larger marginal contribution when worker j's data is absent. By (54) and $A^{-jl} \leqslant A^{-l}$, both (52) and (53) are non-positive: employing worker j at time 1 weakly lowers every other worker's time-2 wage.

Substituting (49) and the non-positivity of (52) and (53) into (47) yields $S_j \geqslant \theta^j > 0$. Full employment at time 1 is therefore consistent with Nash-in-Nash bargaining, which completes the proof. □

B.3 Proof of Theorem 4.13

Proof. We restrict attention to the equilibrium of Proposition 4.11, with full employment and full contributions $k_1 = \theta$. By (32) and (31), we have

$$-\alpha^l(0, \theta^{-j}) \geqslant -\tilde{\alpha}^l(0, \theta^{-j}) \text{ for any workers } j \text{ and } l. \tag{55}$$

By Nash-in-Nash bargaining at time 2 and (27), worker j's time-2 wage under $\left(\alpha^l\right)_{l \in J}$ is

$$\begin{aligned} &\gamma \left(\max\left\{\theta^j, \alpha^j\left(\theta\right)\right\} - \alpha^j\left(0, \theta^{-j}\right) + \sum_{l \neq j} \left(\max\left\{\theta^l, \alpha^l\left(\theta\right)\right\} - \max\left\{\theta^l, \alpha^l\left(0, \theta^{-j}\right)\right\} \right) \right) \\ &\quad \geqslant \gamma \left(\max\left\{\theta^j, \tilde{\alpha}^j\left(\theta\right)\right\} - \tilde{\alpha}^j\left(0, \theta^{-j}\right) + \sum_{l \neq j} \left(\max\left\{\theta^l, \tilde{\alpha}^l\left(\theta\right)\right\} - \max\left\{\theta^l, \tilde{\alpha}^l\left(0, \theta^{-j}\right)\right\} \right) \right), \end{aligned} \tag{56}$$

where the inequality is by (32), (55), and $\max\left\{\theta^l, \cdot\right\}$ non-decreasing. The right-hand side is worker j's time-2 wage under $\left(\tilde{\alpha}^l\right)_{l \in J}$. □

B.4 Proof of Theorem 4.14

Proof. We are comparing full-contribution equilibria, so workers incur no withholding costs. Thus, it suffices to show that workers' wages are lower under individual data ownership than under no surveillance.

There are at least two workers, they have identical maximum contributions, and their contributions are perfect substitutes, so expression (27) for pairwise surplus at time 2 under individual ownership reduces to

$$\max\left\{\theta^j, \alpha\left(k_1^J\right)\right\} - \alpha\left(k_1^{J \setminus \{j\}}\right) = \left\lfloor \theta^j - f(\theta^j) \right\rfloor_+, \tag{57}$$

which is strictly less than θ^j by $f(0) = 0$, $\theta^j > 0$, and f increasing. Thus wages at time 2 under individual data ownership are strictly lower than $\gamma\theta^j$ by $\gamma \in (0, 1]$, which is the wage under no surveillance (using $\alpha(0) = f(0) = 0$).

Next we consider time-1 wages. There are at least three workers, so deviating to not hire a single worker at time-1 has no effect on time-2 output or on time-2 wages. Thus, the pairwise surplus of hiring worker j at time-1 is equal to θ^j, and time-1 wages are the same under individual ownership and under no surveillance.

We have established that individual ownership results in the same wages at time 1 and strictly lower wages at time 2, which by $\psi > 0$ implies that workers are strictly worse off. □

B.5 Proof of Proposition 4.16

Proof. We verify that the following assessment is an equilibrium: full employment in both periods, knowledge contributions $k_1^j = k_2^j = \theta^j$, bilateral time-2 wages $\gamma \left[\theta^l - \alpha\left(k_1^{\overline{J}_1}\right)\right]_+$, and fee shares given by the own-gains rule, so that each member's time-2 compensation is (34). Each worker's compensation (34) is non-decreasing in k_1^j, by α monotone non-decreasing, and withholding is costly, so $k_1^j = \theta^j$ is a best response. Full employment at time 2 is consistent with tie-breaking in favor of hiring, because each bilateral employment surplus $\left[\theta^l - \alpha\left(k_1^{\overline{J}_1}\right)\right]_+$ is non-negative. It remains to prove that full employment at time 1 is consistent with Nash-in-Nash bargaining; that is, that the pairwise surplus from employing each worker j at time 1 is non-negative.

Fix a worker j. Contributions equal $k_1 = \theta$, so no worker incurs contribution costs, because $c^l(\theta^l) = 0$. Abbreviate $A = \alpha\left(k_1^J\right)$ and $A^{-j} = \alpha\left(k_1^{J\backslash\{j\}}\right)$, and write $\phi^m(\overline{J}_1)$ for (34) with time-1 hired set $\overline{J}_1$. Consider the bilateral negotiation between the firm and worker j at time 1, holding the other workers' time-1 wages and contributions fixed. If worker j is employed, the time-1 hired set is J, and the joint payoff of the firm and worker j is

$$\sum_{l\in J}\theta^l - \sum_{l\in J} w_1^l + w_1^j + \psi\left(\sum_{m\in J}\max\left\{\theta^m, A\right\} - \sum_{m\neq j}\phi^m\left(J\right)\right), \tag{58}$$

where worker j's own time-2 compensation does not appear because it is paid by the firm and received by worker j, and so cancels within the pair. If worker j is not employed, the time-1 hired set is $J\backslash\{j\}$. All workers are still employed at time 2; worker j's data is not in the pool, so worker j receives only the bilateral wage, which again cancels within the pair, while each member $m \neq j$ receives $\phi^m\left(J\backslash\{j\}\right)$. The joint payoff is

$$\sum_{l\neq j}\theta^l - \sum_{l\neq j} w_1^l + \psi\left(\sum_{m\in J}\max\left\{\theta^m, A^{-j}\right\} - \sum_{m\neq j}\phi^m\left(J\backslash\{j\}\right)\right). \tag{59}$$

By (34), $\phi^m(J) - \phi^m(J\backslash\{j\}) = \eta\left(\max\{\theta^m, A\} - \max\left\{\theta^m, A^{-j}\right\}\right)$: each member's compensation moves by exactly η times the output change on their own position. Subtracting (59) from (58), the pairwise surplus is therefore

$$S_j = \theta^j + \psi\left(\left(\max\{\theta^j, A\} - \max\left\{\theta^j, A^{-j}\right\}\right) + (1-\eta)\sum_{m\neq j}\left(\max\{\theta^m, A\} - \max\left\{\theta^m, A^{-j}\right\}\right)\right), \tag{60}$$

and every term of (60) is non-negative, by α monotone non-decreasing and $\eta \leqslant 1$. Hence $S_j \geqslant \theta^j > 0$, so full employment at time 1 is consistent with Nash-in-Nash bargaining, which completes the proof. □

B.6 Proof of Theorem 4.21

Proof. Fix the equilibrium of Proposition 4.16: full employment in both periods, contributions $k_1 = k_2 = \theta$ (so no worker incurs contribution costs), and time-2 compensation (34). As in Section B.5, abbreviate $A = \alpha\left(k_1^J\right)$ and $A^{-j} = \alpha\left(k_1^{J\backslash\{j\}}\right)$, and write $A^0 = \alpha(0)$.

Step 1: The firm's equilibrium payoff. Consider the time-1 negotiation between the firm and worker j, holding the other workers' time-1 wages and contributions fixed. In the event of disagreement, worker j is not paid at time 1 and their data never enters the pool; at time 2 they are employed at the bilateral wage $\gamma\left\lfloor\theta^j - A^{-j}\right\rfloor_+$, so worker j's disagreement payoff is $\psi\gamma\left\lfloor\theta^j - A^{-j}\right\rfloor_+$. In the event of agreement, worker j's payoff is $w_1^j + \psi\phi^j$, where ϕ^j denotes their time-2 compensation (34). Nash-in-Nash bargaining awards worker j their disagreement payoff plus a share γ of the pairwise surplus (60), so the time-1 wage satisfies

$$w_1^j \;=\; \gamma S_j \;+\; \psi\gamma\left\lfloor\theta^j - A^{-j}\right\rfloor_+ \;-\; \psi\phi^j. \tag{61}$$

That is, anticipated time-2 compensation is offset one-for-one in the time-1 wage. Consequently, when we compute the firm's equilibrium payoff—time-1 output net of time-1 wages, plus ψ times time-2 output net of time-2 compensation—the compensation terms $\psi\sum_{j\in J}\phi^j$ cancel:

$$\Pi \;=\; \sum_{j\in J}\theta^j \;-\; \gamma\sum_{j\in J}S_j \;-\; \psi\gamma\sum_{j\in J}\left\lfloor\theta^j - A^{-j}\right\rfloor_+ \;+\; \psi\sum_{m\in J}\max\{\theta^m, A\}. \tag{62}$$

Step 2: Reduction to a per-role inequality. By Observation 4.1, the firm's payoff in the no-surveillance equilibrium is

$$\Pi_0 \;=\; (1-\gamma)\sum_{j\in J}\theta^j \;+\; \psi\sum_{m\in J}\Big(\max\left\{\theta^m, A^0\right\} - \gamma\left\lfloor\theta^m - A^0\right\rfloor_+\Big), \tag{63}$$

since each time-2 position produces $\max\left\{\theta^m, A^0\right\}$ and pays the wage $\gamma\left\lfloor\theta^m - A^0\right\rfloor_+$. Substituting (60) into (62), the time-1 output terms cancel against the θ^j terms of $\gamma\sum_j S_j$, and subtracting (63) yields $\Pi - \Pi_0 = \psi\Delta$, where

$$\begin{aligned}\Delta = &\sum_{m\in J}\Big(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^0\right\} + \gamma\left\lfloor\theta^m - A^0\right\rfloor_+ - \gamma\left\lfloor\theta^m - A^{-m}\right\rfloor_+\Big)\\ &- \gamma\sum_{j\in J}\Big(\max\left\{\theta^j, A\right\} - \max\left\{\theta^j, A^{-j}\right\}\Big) - \gamma(1-\eta)\sum_{j\in J}\sum_{m\neq j}\Big(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\}\Big).\end{aligned} \tag{64}$$

Each difference $\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\}$ is non-negative, because α is monotone non-decreasing. Hence Δ is non-decreasing in η, and it suffices to prove that $\Delta \geqslant 0$ when $\eta = 0$. Setting $\eta = 0$, the two sums over pairwise differences in (64) combine into a sum over all ordered pairs (j, m), which we regroup by role m: we obtain $\Delta\big|_{\eta=0} = \sum_{m\in J} D_m$, where

$$\begin{aligned}D_m \;=\; &\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^0\right\} + \gamma\left\lfloor\theta^m - A^0\right\rfloor_+ - \gamma\left\lfloor\theta^m - A^{-m}\right\rfloor_+\\ &- \gamma\sum_{j\in J}\Big(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\}\Big).\end{aligned} \tag{65}$$

We claim that $D_m \geqslant 0$ for each m. Since D_m is affine in γ, it suffices to verify the claim at $\gamma = 0$ and $\gamma = 1$. At $\gamma = 0$, we have $D_m = \max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^0\right\} \geqslant 0$, again by monotonicity.

Step 3: The case $\gamma = 1$. At $\gamma = 1$, the claim $D_m \geqslant 0$ is

$$\begin{aligned}&\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^0\right\} + \left\lfloor\theta^m - A^0\right\rfloor_+ - \left\lfloor\theta^m - A^{-m}\right\rfloor_+\\ &\qquad\geqslant \sum_{j\in J}\Big(\max\left\{\theta^m, A\right\} - \max\left\{\theta^m, A^{-j}\right\}\Big).\end{aligned} \tag{66}$$

For any $z \geqslant 0$, we have $\max\left\{\theta^m, z\right\} - \left\lfloor\theta^m - z\right\rfloor_+ = z$, the identity used in the proof of Observation 4.17. Applying this identity with $z = A^0$ on the left-hand side, and with $z = A^{-m}$ after moving

the $j = m$ term of the right-hand side to the left, (66) is equivalent to

$$\sum_{j \neq m} \Big(\max\{\theta^m, A\} - \max\{\theta^m, A^{-j}\} \Big) \;\leqslant\; A^{-m} - A^0. \tag{67}$$

To establish (67), note that for $x \geqslant y$ we have $\max\{\theta^m, x\} - \max\{\theta^m, y\} \leqslant x - y$. Thus

$$\sum_{j \neq m} \Big(\max\{\theta^m, A\} - \max\{\theta^m, A^{-j}\} \Big) \;\leqslant\; \sum_{j \neq m} \big(A - A^{-j}\big) \;\leqslant\; \big(A - A^0\big) - \big(A - A^{-m}\big) \;=\; A^{-m} - A^0, \tag{68}$$

where the second inequality follows from (29) after subtracting the $j = m$ term, which is non-negative by monotonicity. This proves (67). Hence $D_m \geqslant 0$ for each m, so $\Delta \geqslant 0$, and therefore $\Pi \geqslant \Pi_0$. □

B.7 Proof of Theorem 4.22

Proof. Observe that even under costly contributions, the pairwise surplus from employing an additional worker at time 2 is at least zero, so full employment at time 2 is consistent with equilibrium. Given full employment at time 2, worker j's payoff is still as in (28). By $\left|K^j\right|$ finite and the existence of Nash equilibrium in finite games, for any employed set $\overline{J}_1$, there exists a (possibly mixed) profile of time-1 contributions in which each worker's choice of k_1^j maximizes their expected payoff (that is, the expectation of (28)) given $\overline{J}_1$ and the other workers' mixtures over contributions.

Moreover, every contribution in the support of this profile satisfies $k_1^j \geqslant \theta^j$. To see this, observe that the singleton $\{\theta^j\}$ maximizes $w_1^j - c^j(k_1^j)$, because θ^j is the unique minimizer of c^j, and that the expectation of λ_j is non-decreasing in k_1^j, by α monotone non-decreasing. By Topkis's theorem (Milgrom and Shannon, 1994), every maximizer of the expectation of (28) is at least θ^j. Note that this support property does not rely on Condition C, so there is no circularity in what follows.

It remains to show that full employment at time 1 is consistent with the equilibrium we are constructing. Fix a worker j, and fix a realization k_1 of the time-1 contributions, holding the other workers' mixtures fixed. By the support property, $k_1^l \geqslant \theta^l$ for every l, so Condition C applies to the realized profile. From here, the argument is parallel to the proof of Proposition 4.11, with two changes. First, the time-1 term of the pairwise surplus is $k_1^j - c^j(k_1^j)$ rather than θ^j: repeating the

computation that leads to (47) at the realized profile k_1 gives

$$k_1^j - c^j(k_1^j) + \psi\left(\sum_{m\in J}\left(\max\left\{\theta^m, \alpha\left(k_1^J\right)\right\} - \max\left\{\theta^m, \alpha\left(k_1^{J\backslash\{j\}}\right)\right\}\right) - \gamma\sum_{l\neq j}\left(\lambda_l\left(k_1, J\right) - \lambda_l\left(k_1, J\backslash\{j\}\right)\right)\right). \quad (69)$$

Second, because contributions may be random, the pairwise surplus at time 1 is the *expectation* of (69) over the workers' randomization. At each realization, $k_1^j - c^j(k_1^j) \geqslant 0$ by (40), and the arguments leading to (49)–(53) apply verbatim at the realized profile—the partition (48) may vary with the realization—so each realization of (69) is non-negative. The expectation is therefore non-negative as well, so full employment at time 1 is consistent with Nash-in-Nash bargaining, which completes the proof. □

B.8 Proof of Theorem 4.23

Proof. Let κ_1^{-j} be the random variable representing other workers' time-1 contributions under some equilibrium. Observe that for k_1^j played with positive probability, we have

$$k_1^j \in \arg\max_{\hat{k}_1^j} E\left[w_1^j - c^j(\hat{k}_1^j) + \psi\gamma\lambda_j\left(\hat{k}_1^j, \kappa_1^{-j}, \overline{J}_1\right)\right], \quad (70)$$

where λ_j is as defined in (27). Moreover we have

$$\{\theta^j\} = \arg\max_{\hat{k}_1^j} E\left[w_1^j - c^j(\hat{k}_1^j)\right], \quad (71)$$

because θ^j is the unique minimizer of c^j. Since $E\left[\lambda_j\left(\hat{k}_1^j, \kappa_1^{-j}, \overline{J}_1\right)\right]$ is non-decreasing in $\hat{k}_1^j$, the result follows by Topkis's theorem. □

B.9 Proof of Theorem 4.24

Proof. Let $\mathbf{k}_1$ be the random variable representing the time-1 contribution vector in an equilibrium with the data trust and the own-gains rule, so that worker j's time-2 compensation is $\phi^j\left(\mathbf{k}_1, J\right)$ as in (34). For any worker j, they cannot profitably deviate to contribute θ^j at time 1, so for any k_1^j

in the support of $\mathbf{k}_1$ we have

$$E\left[-c^j(k_1^j)+\psi\,\phi^j\left(\left(k_1^j,\mathbf{k}_1^{-j}\right),J\right)\right]\geqslant E\left[\psi\,\phi^j\left(\left(\theta^j,\mathbf{k}_1^{-j}\right),J\right)\right]. \tag{72}$$

The terms of (34) that do not involve $\alpha(k_1)$ are common to both sides of (72), and by α monotone non-decreasing we have $\max\left\{\theta^j,\alpha\left(\theta^j,\mathbf{k}_1^{-j}\right)\right\}\geqslant\max\left\{\theta^j,\alpha(0)\right\}$, so (72) implies

$$\begin{aligned}E\left[-c^j(k_1^j)\right]&\geqslant-\eta\psi\,E\left[\max\left\{\theta^j,\alpha\left(\mathbf{k}_1\right)\right\}-\max\left\{\theta^j,\alpha(0)\right\}\right]\\&\geqslant-\psi\,E\left[\max\left\{\theta^j,\alpha\left(\mathbf{k}_1\right)\right\}-\max\left\{\theta^j,\alpha(0)\right\}\right],\end{aligned} \tag{73}$$

where the last inequality is by $\eta\in[0,1]$ and the bracketed expectation non-negative. By the same argument as in Section B.8, if k_1^j is in the support of $\mathbf{k}_1$, then $k_1^j\geqslant\theta^j$, so $E\left[\mathbf{k}_1^j\right]\geqslant\theta^j$. Summing (73) across j and adding $E\left[\mathbf{k}_1^j\right]\geqslant\theta^j$ for each j yields

$$\sum_j E\left[\mathbf{k}_1^j-c^j(\mathbf{k}_1^j)+\psi\max\left\{\theta^j,\alpha\left(\mathbf{k}_1\right)\right\}\right]\geqslant\sum_j\left(\theta^j+\psi\max\left\{\theta^j,\alpha(0)\right\}\right). \tag{74}$$

The left-hand side of (74) is the expected total surplus under the data trust, and the right-hand side is the total surplus in the no-AI case, in which each position produces θ^j at time 1 and $\max\left\{\theta^j,\alpha(0)\right\}$ at time 2. □