Canaries in the Gold Mine: Early Productivity Gains from Artificial Intelligence Creating Organization Capital

Tania Babina[a], Alex He[b], Renhao Jiang[c]

[a]*University of Maryland, College Park, NBER, and CEPR*
[b]*University of Maryland, College Park*
[c]*University of Maryland, College Park*

Abstract

Using a new firm-level measure of AI investment based on AI-skilled employment—spanning machine learning through generative and agentic AI—we show that AI investments are associated with productivity growth in recent years, but not over the previous decade. We trace the productivity gains to the accumulation of organization capital that AI helps create: durable firm-specific knowledge acquired through learning-by-doing that enables more efficient production. We build a novel measure of organization capital based on workers' job descriptions and document that productivity gains are driven by AI-skilled jobs that build organization capital. Overall, our findings suggest that AI investment generates productivity growth by creating organization capital.

Keywords: Artificial Intelligence, Productivity, Organization Capital, AI Investments, Intangible Capital
JEL: D22, D24, E22, J24, L23, O33

⋆First draft: February 26, 2026. This draft: August 22, 2026. Emails: tbabina@umd.edu (Tania Babina), axhe@umd.edu (Alex He), rjiang07@umd.edu (Renhao Jiang). We are grateful for discussions and helpful comments from Katharine Abraham, Daron Acemoglu, David Autor, Martin Beraja, Nicholas Bloom, Lee Branstetter, Timothy Bresnahan, Erik Brynjolfsson, David Byrne, Nicolas Crouzet, David Deming, Joshua Gans, Avi Goldfarb, Daniel Gross, Kyle Herkenhoff, Anders Humlum, Xavier Jaravel, Benjamin Jones, Charles Jones, Sophia Kazinnik, Kai Li, Joseba Martinez, Daniel Rock, Raffaella Sadun, Lawrence Schmidt, Gregor Schubert, Bryan Seegmiller, Chad Syverson, Neil Thompson, John Van Reenen, Michal Zator, Miao Ben Zhang, and seminar participants at the Baruch AI in Finance Workshop, Drexel University, the OECD Virtual Conference on the Competition in the AI Landscape, the UBS New Frontiers in Quant Finance Conference, the University of Maryland, the University of North Carolina Alumni Conference, the University of Sydney, the University of Technology Sydney, the UBC Summer Finance Conference, and the Workshop on Emerging Research on AI, Labor, and the Future of Work at the Economic Innovation Group. We thank Yueran Ma for generously sharing the code used to match firms in Revelio to Compustat. We thank Yaru Su for excellent research assistance. All errors are our own.

Technological change is a major source of economic growth and productivity improvements (Aghion and Howitt, 1992; Kogan et al., 2017), yet evidence on the productivity effects of Artificial Intelligence (AI) remains mixed. Micro-level studies of adopting AI tools for specific tasks typically find large productivity gains, while evidence from more aggregated data across AI use cases, firms, or sectors has yet to show consistent productivity improvements. The debate echoes Solow's famous observation about an earlier technological revolution: *"You can see the computer age everywhere but in the productivity statistics."* We revisit this question using new firm-level measures of AI investment spanning the early machine learning wave and the recent generative and agentic AI wave. We ask not only whether productivity gains have arrived, but also what enables firms to realize these gains.

We document that firms investing more in AI experience significantly faster productivity growth from 2018 to 2024, in sharp contrast to the absence of such gains over 2010–2018. These gains are not confined to recent advances in generative AI but extend across a broad range of AI technologies. We trace these gains to organization capital (OC): durable, firm-specific knowledge that enables more efficient production. Using a new measure of organization-capital-building jobs, we show that productivity gains are concentrated in AI investments that build OC, whose gradual accumulation helps explain why AI's productivity benefits emerge only with a long lag. Our key contribution is to show that the productivity gains from firms' AI investments come from organization capital that AI helps create.

Measuring firm-level AI investment is challenging because firms do not systematically disclose it. We therefore extend the human-capital approach of Babina et al. (2024), using a firm's employment of AI-skilled workers as a proxy for its internal AI development and deployment.[1] We combine the near-universe of U.S. online job postings from Lightcast with Revelio Labs resume data covering two-thirds of the U.S. workforce. In job postings, we identify AI-relevant skills based on their co-occurrence with core AI skills such as machine learning and artificial intelligence. Because this procedure is data-driven, it adapts to evolving AI terminology: for example, our AI-relatedness score (ranging from 0 to 1) for *agentic systems* is 0.98, despite the term emerging only recently. The measure therefore spans successive waves of AI, from early machine learning to recent advances in generative and agentic AI.

We then use the identified AI-related terms to classify AI-related jobs in worker resumes, yielding a false-positive rate of only about 2%. The resulting measure captures workers primarily involved in firms' internal AI development, deployment, and management; it

[1] Chequer et al. (2026) link Census surveys to worker tax records and find that about half of firms using AI intensively conduct in-house AI R&D and that talent is a key input to adoption, suggesting that successful AI deployment requires internal investment and specialized AI talent, not just off-the-shelf AI tools.

does not capture firms that rely exclusively on off-the-shelf external AI tools. This also distinguishes our measure from AI labor-exposure measures, which capture the potential for AI to perform workers' tasks. Our approach instead allows us to estimate net effects on firm outcomes across all internal AI applications, not just labor-replacing ones (Babina, 2026). We aggregate AI employment to the firm-year level and link it to U.S. public firms.

The data reveal rapid and broad-based growth in AI investment. By 2024, U.S. public firms employ more than 300,000 AI-skilled workers, representing 2.1% of their U.S. workforce, with AI shares rising in every industry. The composition also shifts over time: employment in earlier machine-learning, NLP, and computer-vision technologies decelerates, while generative and agentic AI accelerates late in the sample. AI investment rates outside U.S. public firms are roughly half as high, suggesting that public firms are ahead in AI investment.

Using a long-differences specification, we relate changes in firms' productivity to changes in AI investments: using *changes* helps address the core concern that AI-investing firms are more productive to begin with. Our 15-year panel of firm AI investments lets us trace how this relationship evolves. Three facts emerge. First, consistent with Babina et al. (2024), AI investment is unrelated to productivity growth during the rise of AI adoption from 2010 to 2018. Second, the relationship turns strongly positive over 2018–2024: a one-standard-deviation increase in the growth of a firm's AI worker share is associated with 6.2 percentage points greater sales-per-worker growth and 7.5 percentage points greater revenue total factor productivity (TFP) growth over six years—roughly one percentage point of additional productivity growth per year. Third, productivity growth is positively related to both earlier and later AI waves, indicating that the effect is not driven solely by generative and agentic AI.

The timing points to substantially lagged and gradually accumulating productivity gains from AI investment. Firms that increase AI investment over 2018–2024 show no differential productivity growth in earlier periods, while estimated effects rise monotonically after 2018. Regressions of ex-post productivity growth on ex-ante AI investment, and vice versa, reveal long lags and confirm that AI hiring precedes productivity gains rather than the reverse.

The results are robust to alternative measures of AI and productivity, and are not explained by acquisitions, offshoring, technology firms, retrospective resume relabeling, pre-existing productivity trends, industry or pandemic shocks, job remote-suitability, or investment in non-AI IT. To address remaining concerns, we instrument firms' AI investments using their pre-2010 hiring ties to universities historically strong in AI research—ties formed before AI's commercial rise (Babina et al., 2024). The IV estimates confirm a significant recent productivity increase among AI-investing firms, supporting a causal

interpretation.

We next ask whether the productivity gains come from producing the same sales with fewer workers or more sales with similar employment. In the earlier period, AI-investing firms expand sales and employment at similar rates. Over 2018–2024, however, AI investment predicts substantial sales growth but no employment growth, indicating a shift toward productivity-driven rather than labor-intensive expansion.

We then turn to mechanisms. General-purpose technologies raise productivity not only by deepening physical capital, but also through complementary intangible investment in organization, skills, and redesigned processes (Bresnahan and Trajtenberg, 1995; Brynjolfsson et al., 2002, 2021). We show that AI-investing firms expand intangibles (e.g., SG&A and R&D) rather than tangibles (PP&E). The intangible buildup is mirrored in the workforce: AI-investing firms shift toward workers with advanced degrees and senior roles, while shares of non-college and junior workers decline. Unlike earlier technologies that spurred capital deepening, AI adoption manifests as intangible accumulation.

Among intangibles, we show that organization capital (OC)—built by AI-skilled workers—is a central mechanism driving productivity growth. OC is durable, firm-specific knowledge accumulated through learning-by-doing that allows firms to convert inputs into output more efficiently (Prescott and Visscher, 1980; Atkeson and Kehoe, 2005; Eisfeldt and Papanikolaou, 2013). AI can accelerate this process because it is fundamentally a learning technology (Turing, 1950; Agrawal et al., 2022): models trained on a firm's own data and embedded in its systems and workflows codify knowledge generated through operations and, in doing so, become part of the firm's OC. Firms' internal AI workers build and maintain these systems in collaboration with employees who possess domain expertise. This process converts knowledge that might otherwise diffuse slowly through experience, mentoring, or documentation into a durable organizational asset.

We formalize this mechanism in a model extending Atkeson and Kehoe (2005): OC is a knowledge stock that accumulates as a by-product of operations, and AI is a *codification* technology that raises the rate at which firms learn from their operations to build OC. Two predictions follow. First, the effect of AI investment is slow to arrive because OC is a stock and AI changes only the flow into it: even though AI helps firms build OC faster, the productivity gains emerge gradually with long lags. Second, the effects should be larger for firms that begin with little OC relative to their size, since firms that have already codified much of what they do have less left to codify.

To test this channel, we develop a new measure of OC-building jobs, motivated by the fact that OC is built by workers and embodied in routines they establish. Using job titles and descriptions, we score whether a job builds OC—durable firm-specific knowledge, systems, processes, or platforms. For example, an employee training and maintaining

large language models on customer-shopping data to improve the firm's sales processes is OC-building, whereas a machine learning engineer building a model for a single client engagement is not.

Our novel OC measure reveals several stylized facts about the organization-capital content of work. First, AI jobs are almost twice as likely as non-AI jobs to build OC (58% vs. 31%). Second, the OC-building share of AI jobs rises from about one-half in 2010 to almost two-thirds in 2024, while the corresponding share of non-AI jobs remains flat at about one-third, indicating a genuine change in the nature of AI work rather than a broader trend of jobs becoming more OC-building. Third, newer AI technologies are more OC-intensive: 66% of generative and agentic AI jobs build OC versus 57% of ML, NLP, and CV jobs. Fourth, AI jobs are more likely to build OC in non-tech firms than in tech firms, consistent with non-tech firms using AI primarily to build internal capabilities rather than products for sale.

The granularity of the resume data lets us describe *how* AI workers build OC. We document a wide range of OC assets AI workers build: they codify proprietary knowledge into reusable models, systems, and platforms spanning product development, marketing, infrastructure, risk, customer support, and scientific work. Essentially every OC category built by AI workers grows between 2010 and 2024, indicating broad-based accumulation of organizational assets. Across eight firm case studies, a common pattern emerges: models are trained on records the firm already produces, turning documentation of past work into a reusable organizational asset. For example, Morgan Stanley trains an AI assistant from its own advisor meeting notes; P&G constructs a digital twin from plant telemetry; and Expedia turns its engineers' incident histories into agentic workflows.

We then show directly that OC links AI investment to productivity. First, growth in OC-building AI jobs predicts productivity growth, whereas growth in other AI jobs does not. Second, AI investment increases firms' measurable stock of OC, using the standard capitalized-SG&A proxy. AI-investing firms increase OC over 2018–2024 but not 2010–2018, matching the timing of the productivity results. This buildup is driven by growth in OC-building AI jobs, mirroring the productivity decomposition. Third, AI-driven growth in both productivity and OC is larger among firms starting with relatively little OC, as the model predicts. Finally, two forces help explain why the productivity effects are stronger in the later period. The share of AI jobs devoted to OC-building has risen, so AI investment is now more often directed toward building OC. And because OC must accumulate before raising productivity, even earlier AI investment pays off only later.

We also examine, and find no support for, two alternative explanations for our findings. If measured productivity rose mainly through labor replacement, AI-investing firms should reduce employment, particularly among highly skilled AI-exposed workers. Instead,

overall employment is unchanged and high-skilled employment rises. If higher revenue productivity simply reflected price increases, AI-intensive industries should show rising prices. They do not.

Using a long panel of firms across industries and AI use cases, we show that productivity gains from firms' AI investments have begun to emerge—roughly a decade after AI's commercial rise. The early adopters in our sample thus serve as canaries—not in a coal mine, warning of danger, but in a gold mine, offering an early signal of the gains that broader AI diffusion may bring. We trace these delayed productivity gains to AI-driven accumulation of organization capital. What distinguishes AI from earlier general-purpose technologies is that organization capital is not only built around AI; AI itself helps firms build organization capital.

Related Literature. Our first contribution is to introduce a new firm-level measure of AI investment spanning earlier and recent waves of AI technologies.[2] It complements measures of labor exposure to AI (e.g., Felten et al., 2018; Webb, 2020; Eloundou et al., 2024; Jiang et al., 2025; Eisfeldt et al., 2026b), workers' use of external AI tools (e.g., Blank et al., 2025; Humlum and Vestergaard, 2026; Bick et al., 2026; Iscenko et al., 2026; Chatterji et al., 2026), survey-based measures (e.g., McElheran et al., 2024; Baslandze et al., 2026; Yotzov et al., 2026; Bonney et al., 2026), and financial disclosures (e.g., Cao et al., 2025; Barrios et al., 2025).[3] Closest to our approach, Hampole et al. (2025) construct task-level measures of exposure to AI applications firms actually use. We instead capture firms' internal AI capabilities—irrespective of their labor exposure to AI—and develop new measures spanning successive AI waves, including the recent generative and agentic wave.

Second, we contribute to the literature on AI and firm performance. Babina et al. (2024) find that over 2010–2018, AI-investing firms grow through product innovation without productivity gains. We interpret their findings and ours as successive stages of the same adoption cycle: product-market gains arrive first, while productivity gains emerge slowly as OC accumulates. Other studies of firm valuations and employment emphasize expected labor augmentation or substitution from AI (Rock, 2019; Jiang et al., 2025; Hampole et al., 2025; Eisfeldt et al., 2026b; Blumenfeld et al., 2026), while Acharya et al. (2026) find innovation gains from external AI products.[4] We identify a distinct

[2] Our firm-level AI measures and all relevant materials will be made publicly available upon publication.

[3] Babina (2026) discusses strengths and weaknesses of these different datasets.

[4] In sectors where the value of data and AI is especially high, such as asset management, AI has large benefits. Bonelli (2026) shows that data-driven VC firms better identify startups resembling prior successes. Cao et al. (2024) show that an AI analyst outperforms most human analysts in stock-price forecasts. Chen et al. (2026) find that AI-oriented hedge funds outperformed earlier in the adoption cycle, while Sheng et al. (2026) find higher abnormal returns among hedge funds more reliant on generative AI.

channel: productivity gains from the accumulation of organization capital created by firms' internal AI investment.

This paper also contributes to the debate on whether AI has raised productivity.[5] Micro-oriented task-level experiments focusing on generative AI often find immediate time savings or increased task productivity in specific settings, such as writing and customer support (e.g., Noy and Zhang, 2023; Brynjolfsson et al., 2025b), but also performance losses outside models' capability frontiers (Dell'Acqua et al., 2026). We instead study firm-wide productivity across a broad range of AI applications and over long horizons.

Firm-level evidence on productivity is more mixed. Survey-based studies have found positive, negative, or no associations between AI adoption and productivity across countries and periods (Czarnitzki et al., 2023; McElheran et al., 2025; Aldasoro et al., 2026; Yotzov et al., 2026; Baslandze et al., 2026). Relative to this work, we contribute in three ways. First, we measure firms' internal AI development and implementation directly through their workforce rather than self-reported adoption. Second, our consistent 15-year-long AI panel—unavailable in survey data—reveals an inflection from no measurable productivity relationship before 2018 to a positive one afterward. Third, we document that firms' internal AI investments build organization capital, offering a new explanation for AI's delayed productivity gains.[6]

Finally, we contribute to the literature on organization and intangible capital. First, we develop a new measure of firms' OC-building capabilities based on workers' job descriptions, complementing accounting-based measures of OC (e.g., Lev and Radhakrishnan, 2005; Eisfeldt and Papanikolaou, 2013; Peters and Taylor, 2017) and newer text-based measures constructed from disclosures (Eisfeldt et al., 2026a; Charoenwong et al., 2026), customer capital (He et al., 2024), employee reviews (Cai et al., 2026), and startup job titles (Fahlenbrach et al., 2026).[7] Second, we show that investment in a new technology—AI—can create OC and thereby raise productivity, adding to work on the growing importance of organization and intangible capital (e.g., Lustig et al., 2011; Crouzet et al., 2022; Falato et al., 2022; Crouzet and Eberly, 2023). Third, we extend Atkeson and Kehoe (2005) with a framework in which AI accelerates the rate at which learning-by-doing is codified into

[5] See Imas (2026) for a discussion of this debate, and Bryan (2026) and Babina (2026) for broader discussions of AI's economic impacts.

[6] This delay is usually attributed to complementary co-investments, adjustment costs, and organizational change (Brynjolfsson et al., 2019); Baily et al. (2025) reach a similar conclusion from the history of general-purpose technologies.

[7] More broadly, our paper adds to a literature using textual data to measure firm attributes. Seminal work extracts information from corporate filings (e.g., Hoberg and Phillips, 2016); see Gentzkow et al. (2019) for a survey. A recent wave of papers uses LLMs to extract signals from corporate texts (Breitung and Müller, 2025; Acikalin et al., 2026; Li et al., 2026). Closest to our setting is the literature applying text analysis to labor market data (Deming and Kahn, 2018; Campello et al., 2024; Kalyani et al., 2025).

organization capital.

1. Data

1.1. Data on Job Postings from Lightcast

We use U.S. job postings data from Lightcast over 2010–2025. The data contain approximately 456 million postings scraped from online job boards and company websites, standardized and deduplicated by Lightcast (Lightcast, 2026a). The dataset covers a near-universe of U.S. online vacancies (Lightcast, 2026b) and has stable occupational representativeness over time (Hershbein and Kahn, 2018).

Most importantly for our purposes, each job posting includes standardized skill requirements, which we use to identify AI-relevant skills. We retain non-internship postings with an identified employer and at least one standardized skill, yielding a final sample of 367 million postings covering 32,309 unique skills.

1.2. Data on Employment Profiles from Revelio Labs

We obtain worker employment histories from Revelio Labs in March 2026. The data provider compiles publicly available resumes from professional networking websites (e.g., LinkedIn).[8] For each resume, we observe job start and end dates, job titles and descriptions, employer names, employee locations, and educational background, including universities and majors. Revelio uses machine-learning methods to assign seniority levels and occupational classifications to each job spell. Revelio imputes worker salaries based on worker and job characteristics, including job title, years of experience, seniority, employer, and location. Although online professional profiles tend to over-represent white-collar workers, the data cover approximately two-thirds of U.S. workforce and provide a broadly representative breakdown by industry (Revelio Labs, 2026). The detailed job-spell information yields a job-worker-firm-year panel and allows us to construct consistent firm-year workforce measures.

We dynamically match Revelio employers to Compustat firms using subsidiary information and firm-identifier crosswalks, accounting for acquisitions and divestitures. Appendix A.1 describes the linking procedure in detail and provides an extensive analysis of Revelio's coverage of U.S. public firms' employees. We end the Revelio–Compustat matched firm panel in 2024 for two reasons. First, as of the Revelio data download date,

[8]While Babina et al. (2024) use Cognism as a source of resume data, we use Revelio Labs in this paper. Both datasets draw on online resumes and provide comparable information on workers' employment histories and characteristics, allowing us to construct measures conceptually consistent with those in Babina et al. (2024). The added advantage of Revelio data is its extended time series coverage (Cognism data end in 2021).

many U.S. public firms had not yet reported fiscal 2025 financials, which substantially reduces the sample in 2025. Second, workers update their resumes with a lag, creating a well-documented source of measurement error (Tambe et al., 2020; Babina et al., 2024). Of the 1.07 billion U.S.-based person-firm-year employment records between 2010 and 2024, 211 million (20%) are matched to U.S. public firms (rather than private firms or public-sector employers), covering 30.5 million distinct employees. This match rate to public firms is consistent with: i) Revelio data covering approximately two-thirds of the U.S. workforce, and ii) publicly listed U.S. firms accounting for approximately 26% of the overall U.S. employment (Davis et al., 2006). For our focal 2018–2024 period, the sample contains 111 million person-firm-years and 23.4 million distinct employees.

1.3. Additional Data Sources

Firm-level variables, including sales, assets, employment, and SG&A, come from Compustat. We obtain commuting-zone wages and education from the American Community Survey, industry wages and employment from the Quarterly Workforce Indicators, and producer price indices from the Bureau of Labor Statistics (BLS).

2. Measurement and Descriptive Evidence

2.1. Measuring and Validating Firms' Internal AI Investments

We build on and extend Babina et al. (2024) to measure firms' internal AI investment using their employment of AI-skilled workers. AI-skilled labor is a key input into firms' AI development and implementation, while other inputs such as data and compute are complementary to this labor. Our human-capital-based measure therefore captures the relative intensity of firms' internal AI investments. Appendix A.2 provides additional construction and validation details.

Our research question—how firms' productivity changes as they invest in internal AI capabilities—requires data that capture internal AI development and implementation across a broad set of firms and use cases, and that reflect actual AI capabilities rather than perceptions of them.[9] Our methodology is reproducible, inexpensive to implement at scale, and data-driven, allowing it to capture evolving AI job terms across successive waves of AI. It is not intended to capture workers who use AI tools for productivity without working on internal AI implementation (e.g., chatbot users), and we show below that such workers are rare in our sample.

[9]See Babina (2026) for a discussion of the requirements for measuring firms' internal AI capabilities, including broad firm and use-case coverage, internal rather than outsourced development, and actual rather than perceived capabilities.

Our methodology has four steps: (i) identify AI-relevant skills in job postings; (ii) expand these skills into terms that may appear in resumes; (iii) screen candidate search terms for false positives; and (iv) search for the validated terms in workers' job titles and descriptions to identify AI-related workers. This data-driven procedure avoids relying on a fixed, pre-specified AI keyword list that may miss new terminology or include ambiguous terms.

Step 1: AI-relevant skills from job postings. We look at the required skills listed in job postings and measure their co-occurrence with four core AI skills: *artificial intelligence*, *machine learning*, *natural language processing*, and *computer vision*.[10] Specifically, for each required skill s, we compute the following measure:

$$w_s^{AI} = \frac{\text{\# of job postings requiring skill } s \text{ and (ML, NLP, CV, or AI in required skills)}}{\text{\# of job postings requiring skill } s} \quad (1)$$

This score captures how closely a skill is associated with the core AI technologies. For instance, *Agentic Systems* has a score of 0.98, meaning that 98% of postings requiring it also list at least one core AI skill, and *Large Language Modeling* has a score of 0.93. By contrast, *Microsoft Office* has a score of only 0.003. Appendix Table A.6 reports the 50 skills with the highest AI-relatedness scores and confirms that the methodology captures both established AI concepts (e.g., *Doc2vec*) and concepts that have emerged more recently (e.g., *AgentGPT*). We retain skills with scores of at least 0.6 that appear in at least 50 postings, yielding 335 candidate AI-relevant skills.[11]

Step 2: Expansion into resume search terms. Because resume text is unstructured and workers may describe the same skill using abbreviations or alternative spellings, we expand each seed skill into plausible textual variants. This step yields 433 unique search terms. We then augment this list using an LLM to generate synonyms, alternative spellings, and closely related terms (like *TFLite* and *TF-Serving* for *TensorFlow*). Each proposed term is then manually validated using web-search results before it enters the candidate dictionary. This step adds 89 search terms.

Step 3: False-positive screen. We inspect candidate search terms for common non-AI meanings and remove those with high false-positive rates—for example, *RL* often refers to Ralph Lauren rather than reinforcement learning. The final AI dictionary contains 510 search terms.

Step 4: Identifying AI-related jobs and constructing firm-level measures. We classify a job as

[10] We experimented with adding newer AI skills, such as agentic and generative AI, to the core AI skill list, but these yield few additional search terms and risk introducing look-ahead bias.

[11] Lowering the AI-relatedness threshold from 0.6 to 0.5 mostly adds non-AI-specific terms; the few AI-specific ones capture almost no additional jobs.

AI-related if its title or description contains at least one of the 510 validated search terms. For our baseline analysis of U.S. public firms, we measure AI investment as the share of a firm's U.S.-based jobs classified as AI-related in each year. For broader descriptive analyses, we apply the same classification to all U.S.-based workers, regardless of employer type.

We also introduce AI measures by AI technology subcategory. We assign each AI keyword to one of seven categories: general AI, machine learning, natural language processing, computer vision, generative AI, agentic AI, and other. A job is classified into a category if its title or description contains at least one of that category's keywords. Categories are not mutually exclusive because a job may use multiple technologies. We further define Wave I jobs as those matching machine learning, NLP, or computer vision terms and Wave II jobs as those matching generative or agentic AI terms.[12] Appendix A.2.4 provides the keyword classification and job examples in each subcategory. While AI use cases are heterogeneous, a common pattern is that a worker uses AI to learn about some aspect of a firm's business and improve it. For example, a "machine learning" job builds a random forest model identifying drivers of customer satisfaction to guide resource allocation and marketing; a "computer vision" job automates refund approval workflows, cutting processing time and error; and an "agentic AI" job integrates LLM summarization to generate investment insights and market summaries.

As a starting point for validating our AI measures, we find a low false-positive rate of around 2% for AI-related jobs. More generally—in the absence of comprehensive data on firms' AI investments—we validate our AI measures by examining the characteristics of jobs that our methodology identifies as AI-related.[13] Consistent with AI workers being highly skilled (Gofman and Jin, 2024) and highly paid (Akcigit et al., 2026), we find that they are substantially more educated, more likely to hold senior roles, and more highly paid (Appendix Figures A.7, A.8, and A.9).

Appendix Tables A.13 through A.16 list the job titles and occupations that account for the most AI jobs. They make clear that titles and occupations alone are a poor way to identify AI-skilled talent: "Software Developers" is the second most common occupation containing AI jobs, yet only 5.5% of Software Developers are tagged as AI-related. Employee job descriptions instead provide richer context for identifying AI-related jobs.

We next classify AI jobs into functional categories (see Appendix A.3). Appendix

[12]For analyses requiring mutually exclusive groups, we define three groups: jobs that match Wave I but not Wave II, Wave II but not Wave I, and both.

[13]Validating data on internal AI investments against other firm-level AI data is conceptually problematic (Babina, 2026): AI patents capture AI *outputs* by AI inventors rather than the *inputs* and downstream use we study; disclosure-based measures reflect strategic release decisions and investor demand for information; and surveys depend on design, response rates, and respondent qualifications, and rarely have a historical panel structure.

Tables A.24 and A.25 show that about 75% of AI jobs involve the internal development, implementation, management, coordination, or support of AI efforts within firms: roughly half develop and deploy AI internally, 23% manage, oversee, or govern these efforts, and 7.4% sell and support AI products or services. The remaining 25% contain legitimate AI-related keywords and serve some AI-related function but are not focused exclusively on AI implementation: 17.2% are general digital or technical roles with AI tooling, and 7.5% are incidental or low-confidence mentions. Thus, our resume-based measure primarily captures jobs that contribute to internal AI development and deployment or AI product management and monetization, rather than workers merely using AI tools: only 0.9% of AI-related jobs are classified as *"Non-AI role using AI tools as productivity aid."* Appendix Figure C.5 provides a complementary view using Revelio's seven broad job functions. Most AI jobs are in the "Engineer" and "Scientist" functions, but the share outside these two rose from 25% in 2010 to 33% in 2024 (Appendix Figure C.6), consistent with AI activity shifting from research toward commercialization and deployment across business units.

As a further face-validity check, firms with the highest measured AI employment include well-known AI-intensive firms such as NVIDIA, Amazon, and J.P. Morgan Chase. The firm-level AI series generally increase gradually over time, consistent with AI investment being a slow-moving process, as documented by Babina et al. (2024) (Appendix Tables A.17–A.22 and Figures A.11–A.12).

We further validate our new AI measure against two independent data sources in Appendix A.2.8 and A.2.9. First, over 2010–2021, our Revelio-based firm-level AI job counts correlate 70–90% across years with the Cognism-based measure in Babina et al. (2024). Second, firms' new AI hires in Revelio closely track their advertised demand for AI-skilled labor in Lightcast job postings, with a within-firm-year correlation of 0.89. These high correlations with historical resume and job postings datasets also alleviate concerns that workers retroactively relabel past positions by adding AI terminology that became popular only later (Bloom et al., 2026).

2.2. *Measures of Organization Capital*

A central mechanism we examine is that AI investments raise productivity by helping firms build organization capital (OC)—durable, intangible, firm-specific knowledge acquired through learning-by-doing that enables firms to operate more efficiently. We therefore construct two complementary firm-level measures of OC.

Our first measure is a standard accounting-based stock of OC, constructed by capitalizing a firm's past selling, general, and administrative expenses (SG&A) (Eisfeldt and Papanikolaou, 2013). We apply the perpetual inventory method to deflated SG&A expenses using a 20% annual depreciation rate, and then scale the resulting OC stock by

property, plant, and equipment (PP&E) or total assets.[14]

Our second measure introduces a novel method to identify jobs that build OC based on worker job descriptions (see Appendix B.2 for details). Because OC is partly built by workers and embodied in firms' routines, systems, and processes, worker job descriptions—which detail the tasks workers perform, the routines, systems, and processes they build and maintain, and the firm-specific knowledge they develop to improve operations—provide direct evidence on whether a worker is creating durable firm-specific capabilities. We classify all AI jobs and a random sample of non-AI jobs according to whether the job contributes to building OC. We score each job in both samples with an LLM that reads the worker's job title and description and returns an OC score between 0 and 1. We classify a job as OC-building if its score is at least 0.6, the threshold between mixed and substantial evidence of building OC. For AI-related jobs, we aggregate this job-level classification to the firm-year level by computing two shares of total jobs—those that are related and unrelated to building OC.[15]

We further categorize *how* workers build OC by classifying jobs into standardized OC categories. The LLM first identifies recurring types of OC-building activity from the job descriptions, yielding 18 categories, and then assigns each job to the closest category. Appendix Table B.5 reports the resulting taxonomy and definitions.

In the absence of granular OC data, we validate our new measure in several steps. First, Appendix Tables B.9 and B.10 provide job examples and show that our OC classification requires concrete evidence of building durable organizational capabilities. Appendix Table B.9 contrasts AI job descriptions that do and do not build OC and shows that the presence or more frequent mentions of AI terms does not by itself constitute OC building. Examples of OC-building AI jobs include training and maintaining LLMs on customer shopping behavior to improve sales processes, or an AI product manager redesigning business processes and workflows to improve productivity. In contrast, a machine learning engineer building a model for a single client engagement, or a data scientist running a one-off analysis to answer a specific business question, is not OC-building: both do internal technical work, but the output is a model or an analysis that may never be embedded in the firm's systems and workflows. Overall, OC-building AI jobs create systems the firm continues to use, while other AI jobs apply similar technical skills to one-off outputs, portable methods, or client work.

[14] We find similar results using a 15% depreciation rate.

[15] One concern is that our OC classification simply picks up workers who write higher-quality job descriptions. Appendix B.2 shows this is not the case: OC-building jobs have longer descriptions, but they use more repetitive vocabulary, cite fewer specific numbers and tools, and contain spelling errors at similar rates, and among descriptions of similar length the two groups are indistinguishable across all writing quality measures.

Second, word clouds of job descriptions (Appendix Figure B.9) show that among AI jobs, both OC-building and non-OC-building jobs share *data*, *ai*, and *model*, but OC-building roles emphasize *team*, *lead*, *product*, *business*, *design*, and *process*, indicating team leadership and business process design, whereas non-OC-building jobs mention *develop*, *machine*, and *learning*, emphasizing technical model development. Among non-AI jobs, OC roles emphasize *team*, *process*, *lead*, *management*, and *business*—operational and managerial routines—versus *customer*, *service*, *sale*, and *product* for non-OC roles.

Third, we also examine OC-building AI jobs of eight firms across industries in detail in Appendix B.3. In each case, the jobs our measure flags as OC-building are the ones tied to systems that outlast any single project—models trained on the firm's own research documents, meeting transcripts, transactions, sensor readings, or product data—confirming that the classification picks up durable firm-specific assets rather than generic technical work. Section 4 returns to these case studies to illustrate how AI creates OC.

Fourth, the classification also varies intuitively across job functions: 79% of AI-related workers in management roles are classified as OC-building (Lustig et al., 2011; Dessein and Prat, 2022), compared with 22.6% in support for AI products and services—consistent with customer-service roles serving clients rather than building OC. Only about half of workers in core AI-development roles are classified as OC-building. These patterns also hold using standard occupations provided by Revelio (Appendix Tables B.6–B.8). Thus, our OC classification does not simply equate technical or internal AI work with OC-building; it isolates jobs that genuinely build durable firm-specific capabilities.

Finally, Section 4.2 provides external validation: firms with more OC-building AI jobs subsequently accumulate more OC as measured by capitalized SG&A.

2.3. Measures of Productivity

We use two main productivity outcomes: log sales per worker and revenue TFP. We construct revenue TFP as the residual from industry-specific regressions of log real sales on log employment and log capital, including firm and year fixed effects. Revenue TFP assumes a Cobb–Douglas production function, whereas sales per worker provides a measure of labor productivity that does not require this restriction. As a robustness measure, we also construct TFP using the control-function approach (Olley and Pakes, 1996), following the implementation in Imrohoroğlu and Tüzel (2014) (see Appendix A.4).

2.4. Descriptive Evidence

AI Measures. Most of our descriptive results and our main regression sample use U.S.-based employees of U.S. public firms. Figure 1 shows that AI employment rises sharply over 2010–2024, reaching more than 300,000 jobs, or 2.1% of employment, in 2024. Because

AI workers are highly paid, the wage-weighted AI share is larger, reaching 3.3% of firms' wage bills in 2024 (Appendix Figure A.10).

We next examine AI jobs by technology subcategory. Figure 2 shows the most frequent words in job titles and descriptions by AI subcategory. Despite category-specific terms, *data* is common across all categories, consistent with its central role in AI applications. Figures 3 and 4 show a clear shift in AI activity across technology waves. Earlier technologies—machine learning, NLP, and computer vision—decelerate late in the sample, while generative and agentic AI expand rapidly; generative-AI jobs, for example, rise from near zero before 2020 to 0.45% of employment by 2024. The AI waves nevertheless remain closely connected: 58% of agentic AI jobs also contain generative AI keywords, 40% of generative AI jobs mention ML, NLP, or CV keywords, and firms' Wave I and Wave II AI shares have an employment-weighted cross-firm correlation of 0.82 in 2024. This pattern is consistent with the organizational technology ladder in Schubert (2025), in which capabilities built for an earlier technology speed the adoption of the next one.

AI employment grows across all industries, as shown in Figure 5. The Information sector has the highest AI share, rising from 0.7% in 2010 to 5.7% in 2024, but the broad-based increases across sectors are consistent with AI diffusing as a general-purpose technology.

We find similar trends outside U.S. public firms—overall AI job growth, a shift from older toward newer AI technologies, and increases across sectors—but at substantially lower levels. The 2024 AI-job share outside public firms is about 1%, roughly half the 2.1% share at public firms (Appendix Figures C.1–C.3). Even among public firms, AI intensity increases with firm size (Appendix Figure C.4).

Most workers entering their first AI role at a public firm are existing U.S. workers switching out of non-AI roles—roughly 70% throughout the sample. Recent U.S. university graduates are the second-largest and fastest-growing group, doubling from 13% in 2010 to 26% in 2024 (Appendix Figure C.7). These patterns complement Baviskar et al. (2026), who focus on the elite segment of AI talent trained by frontier academic scientists at U.S. universities: beyond elite graduate hiring, firms expand their AI workforce primarily by redeploying workers from non-AI roles.

AI intensity varies substantially across firms, providing the cross-sectional variation used in our analysis. In 2024, the mean firm has an AI-job share of 1.46% with a standard deviation of 3.71%, and the dispersion widens over 2018–2024 (Appendix Table C.1 and Appendix Figure A.6).[16]

[16]The firm-level average AI share in 2024 (1.46%) is lower than the employment-weighted share across all jobs in those firms (2.1%, Figure 1) because larger firms have higher AI shares.

Organization Capital (OC) Measures. Next, we provide stylized facts about our new OC-building job measure. AI jobs are substantially more OC-intensive than non-AI jobs: 58% of AI jobs are classified as OC-building, compared with 31% for non-AI jobs (Appendix Table B.4).[17] This difference grows over time. Figure 6 shows that the OC-building share of AI jobs rises from about one-half in 2010 to almost two-thirds in 2024, while the corresponding share for non-AI jobs remains near one-third. The stability of the non-AI series indicates that the trend reflects a change in the nature of AI work rather than a general shift toward OC-building jobs. Consistent with this, 66% of Wave II jobs are OC-building versus 57% of Wave I jobs, with agentic AI jobs having the highest share at 71% across all AI technologies. We also find that the share of AI workers classified as OC-building is higher among non-tech firms than among tech firms (Appendix Table B.6), consistent with tech firms using AI primarily to build AI products and services for sale, while non-tech firms use AI to build capabilities internally (Appendix Table A.26).

Table 1 shows that AI and non-AI jobs build OC in different ways. AI jobs concentrate in codifying data, models, and proprietary knowledge into reusable technical systems—data platforms (9.1% of AI jobs), proprietary-knowledge generative-AI/NLP (6.9%), reusable AI/ML platforms (4.8%), and firm-specific predictive and decision models (2.3%). Non-AI OC-building jobs are more operational, concentrating in hiring and talent development (5.8%) and manufacturing and process systems (3.5%). Cross-functional operating routines rank among the top categories for both OC groups, highlighting within-firm coordination as a source of OC. AI workers also build a wide range of other OC assets: product and design systems (4.7%), marketing and revenue systems (4.1%), cloud infrastructure and reliability operations (3.4%), financial and risk decision systems (3%), customer support (2%), security and compliance (1.9%), R&D and scientific platforms (1.4%), and others. Among jobs classified as non-OC-building, AI jobs concentrate in portable analytics and one-off models (12.6%), while non-AI jobs are dominated by standardized operations and support (37.9%).

Figure 7 shows that OC-building AI work expands broadly across OC categories between 2010 and 2024 rather than being driven by a single application. Firms accumulate many types of firm-specific assets—data platforms, reusable AI/ML platforms, product and design systems, and cross-functional operating routines. The composition also shifts toward codifying proprietary knowledge, with proprietary-knowledge generative-AI and NLP systems growing from negligible in 2010 to among the largest categories by 2024. This breadth indicates that AI workers build a wide range of organizational assets.

[17] Because the OC classification requires job descriptions, the 58% OC share among AI-related jobs with descriptions corresponds to roughly 50% of all AI jobs in our sample—including those without job descriptions—being classified as OC-building.

3. Main Results: Firm AI Investments and Productivity Growth

3.1. Long-Differences Specification

Firm-level AI investment evolves gradually: AI shares typically rise over many years rather than jumping in any single year. Annual specifications thus contain little identifying variation and may miss productivity effects that materialize only over longer horizons. Therefore, we use a long-differences regression, as is common for slow-moving technological change (Acemoglu and Restrepo, 2020), relating changes in firm productivity to changes in the AI-job share.[18] Differencing removes permanent level differences in productivity between more- and less-AI-intensive firms and focuses the analysis on whether *changes* in firm productivity relate to *changes* in AI investments. For example, for the sample period of 2018–2024, we estimate the following regression:

$$\Delta Y_{i,[2018,2024]} = \alpha + \beta \Delta Share\ AI\ Jobs_{i,[2018,2024]} + Controls'_{i,2018}\gamma + Industry\ FE + \epsilon_i, \quad (2)$$

where the dependent variable, ΔY_i, is the change in firm productivity over the sample period. The main independent variable, $\Delta Share\ AI\ Jobs_i$, is the change in the share of AI jobs at firm i, standardized to have a mean of zero and a standard deviation of one. *Industry FE* denotes 2-digit NAICS industry sector fixed effects. In some specifications, we add a rich set of baseline controls measured at the start of the sample period: firm characteristics (log sales, cash/assets, R&D/sales, markup computed following De Loecker et al. (2020)), and the log Revelio employment count, log industry wage, and local labor-market conditions.[19] To account for differences in the precision of the AI measurement across firms, the regressions are weighted by the number of jobs in Revelio at the start of the period.[20] Standard errors are clustered at the 5-digit NAICS level.

Appendix Table C.2 details the construction of our regression sample.[21] Because our

[18] We use the AI-job share rather than the count because faster-growing firms mechanically add more jobs of all types. AI shares are not negatively correlated with non-AI employment, so the measure does not simply identify firms shrinking their non-AI workforce.

[19] Commuting-zone controls include log average wage, log population, exposure to Chinese import competition (Autor et al., 2013), and the shares of foreign-born, routine, manufacturing, finance, IT-occupation, college-educated, and female workers. When firms span multiple commuting zones, we use the average across commuting zones weighted by the firm's employees in each commuting zone in the Revelio data.

[20] Since the numbers of worker resumes are correlated with firm size, this weighting scheme also roughly weights firms in accordance with their contribution to the economy.

[21] We start with 4,022 Compustat firms that are matched to the Revelio resume data and have a coverage ratio (number of workers in Revelio over Compustat employment) between 0.05 and 3. Among these firms, 3,270 are U.S.-headquartered, 2,748 have at least 50 U.S. jobs in the Revelio data, 2,262 have more than 50% of jobs in the U.S., and 2,215 have Compustat employment of at least 50. Appendix A.1.7 provides extensive analysis highlighting data issues that motivate these data filters. We exclude firms in two 2-digit NAICS tech sectors—51 ("Information") and 54 ("Professional, Scientific, and Technical Services")—and in

focus is the *use* of AI by a wide range of firms rather than the *production* of AI, the baseline sample excludes firms in technology sectors and semiconductor manufacturing.

Our 15-year panel of firm AI investment—rarely available in survey data—lets us trace how the AI–productivity relationship evolves over time. Brynjolfsson et al. (2019) argue that the productivity gains from AI may materialize only with a substantial lag. Consistent with this, Babina et al. (2024) find no correlation between firm AI investments and productivity growth over the early adoption cycle from 2010 to 2018. Panel A of Table 2 confirms this null result: coefficients are small and statistically insignificant for both sales per worker and revenue TFP. Replicating this null result is a meaningful check, since we use a new resume source and a substantially updated classification of AI jobs.

Panel B shows a sharp change after 2018. With the full set of controls, a one-standard-deviation increase in the growth of a firm's AI-job share over 2018–2024 is associated with 6.2 percentage points greater sales-per-worker growth (column 2) and 7.5 percentage points greater revenue-TFP growth (column 4)—roughly one percentage point of additional productivity growth per year. These effects are economically meaningful: they correspond to 0.14 of a standard deviation in sales-per-worker growth and 0.19 of a standard deviation in TFP growth, indicating that changes in the AI-job share account for a substantial portion of the variation in firm productivity growth.

The binned scatterplots in Appendix Figure C.8 show the same pattern nonparametrically: conditional on industry fixed effects and baseline controls, AI investment is unrelated to productivity growth over 2010–2018 but positively related over 2018–2024. Because these plots impose no functional form, they also confirm that this relationship is not an artifact of outliers.

To distinguish whether higher productivity reflects lower employment or higher output, Appendix Table C.4 examines sales and employment growth separately. Over 2010–2018, AI-investing firms grow sales and employment at similar rates, confirming the finding in Babina et al. (2024).[22] Over 2018–2024, however, AI investment predicts higher sales growth but no employment growth: a one-standard-deviation increase in AI investment is associated with 6.7 percentage points greater sales growth, while the employment coefficient is near zero. The employment null result is also present in U.S.-based Revelio data, both for total employment and after excluding AI workers themselves (Appendix

semiconductor manufacturing (NAICS 334413), leaving 1,894 firms. After dropping firms missing 2018 or 2024 data or control variables, the final regression sample contains 1,321 firms. Appendix Table C.3 reports summary statistics for these firms.

[22]The magnitudes are smaller than in Babina et al. (2024) due to differences in the coverage of public firms in Revelio and Cognism resume datasets. In Appendix Table C.5, we show that restricting the sample to firms in Revelio data that also exist in Cognism makes the coefficients larger and closer to the ones in Babina et al. (2024).

Table C.6).[23] Thus, the productivity gains in the later period reflect rising sales with roughly unchanged headcount rather than worker reductions.

The recent productivity gains from AI are also broad-based across industries rather than concentrated in a few sectors: Appendix Table C.7 shows positive effects in Manufacturing, Wholesale and Retail, Finance, Other non-tech industries, and the tech sector (which we exclude from the baseline sample), consistent with AI's promise as a general-purpose technology to raise productivity broadly (Goldfarb et al., 2023).

In Table 3, we examine whether productivity gains are associated with the earlier Wave I AI technologies (ML, NLP, and CV) or the more recent Wave II technologies (generative and agentic AI). We regress productivity growth on the growth in the share of jobs with only Wave I AI skills, only Wave II AI skills, or both. We estimate these components separately because changes in Wave I and Wave II AI jobs are highly correlated (0.77 over 2018–2024), reflecting both technological overlap and the fact that newer AI capabilities often build on earlier ones.

We find that both waves of AI technologies are positively associated with productivity growth over 2018–2024, so the result is not confined to recent generative and agentic AI advances. Because neither wave's search terms include generic terms such as "artificial intelligence," these results also mitigate concerns that the effects are driven by workers adding generic AI terms to their resumes. The positive Wave II effect is notable given the recency of these technologies. Two considerations help explain this finding. First, Wave II jobs are considerably more likely to build organization capital than Wave I jobs, which we show in Section 4.2 is a key mechanism underlying AI-driven productivity growth. Second, the waves are not independent. The data platforms, pipelines, and instrumentation built during the Wave I period provide the raw material on which generative models are trained. Thus, firms that invested early enter the Wave II period with this foundation already in place and can realize productivity gains more quickly than they otherwise would.

3.2. Dynamic Effects

Because firms build AI capabilities gradually and the productivity effects are uncertain and long-lagged, the timing of those effects is difficult to pin down. We exploit the length of our AI investment panel by regressing productivity changes over different windows on firms' 2018–2024 changes in AI-job share. This allows us to examine productivity dynamics both before and during the focal 2018–2024 period. Figure 8 plots the resulting coefficients. For pre-period windows ending in 2018, coefficients are close to zero for both

[23] The Revelio and Compustat employment estimates are nearly identical in magnitude. Because Compustat reports worldwide headcount while we restrict the Revelio measure to U.S.-based positions, this suggests that AI-investing firms are not differentially offshoring employment.

sales per worker and revenue TFP, indicating that firms that later invested in AI were not already on differential productivity trends. For post-2018 windows starting in 2018, the coefficients rise monotonically as the horizon lengthens, consistent with productivity gains from AI accumulating gradually. The pattern is not specific to using 2018 as the base year: Appendix Figure C.9 obtains the same flat pre-period and rising post-period coefficients using 2017 or 2019 as alternative base years.

We next test whether earlier AI investment predicts subsequent productivity growth directly. For each base year t, we regress productivity growth over the subsequent four years on AI-investment growth over the preceding four years—for example, 2018–2022 productivity growth on 2014–2018 AI investment.[24] Appendix Table C.8 shows a positive relationship across recent base years, confirming that AI investment precedes the productivity gains.

We then reverse the ordering to test whether the relationship instead reflects reverse causality: if more productive firms invest more in AI because they have more resources or better infrastructure, past productivity growth should predict future AI investment. Appendix Table C.9 regresses productivity growth from year $t-4$ to t on AI investment from year t to $t+4$, and the coefficients are small and statistically indistinguishable from zero across specifications. Therefore, firms that experienced faster productivity growth in the preceding years did not subsequently invest more in AI.

Together, the timing evidence points to lagged and gradually accumulating productivity gains rather than continuation of pre-existing trends: future AI investments show no differential productivity growth before the investment period but gradually increasing productivity afterward; AI investment predicts subsequent productivity gains; and prior productivity growth does not predict later AI investment. Establishing these patterns requires a panel long enough to observe firms well before and after they invest—something survey data typically cannot offer.

3.3. Robustness

First, Appendix Table C.10 uses the control-function TFP measure of Imrohoroğlu and Tüzel (2014) and yields similar results. We do not use it as our baseline TFP measure because its additional capital and investment data requirements substantially reduce the

[24] We use four-year windows to balance two concerns. Shorter windows risk understating the effect: productivity gains from AI investments take time to realize, so a one- or two-year horizon might capture mostly adjustment costs rather than realized gains. Longer windows attenuate identification for two reasons. First, our data end in 2024, so extending the growth window would force us to drop later base years and shrink the sample. Second, as the horizon lengthens, firms undertake additional rounds of investment and are exposed to more intervening shocks, making it harder to attribute later productivity growth to AI investment in the initial window. Results using three- and five-year windows are similar but less precise.

sample. Re-estimating our baseline productivity measures on the same restricted sample yields comparable coefficients, indicating that the result is not driven by the sample restriction.

Second, the higher productivity observed in the later period does not appear to reflect firm entry, exit, or shifts in sample composition: restricting the sample to firms already present in 2010 yields similar results (Appendix Table C.11), and AI investment does not predict sample exit over 2018–2024.

Third, the results are robust to alternative measures of AI investment. Weighting AI jobs by wages rather than headcount to account for heterogeneity in the quality of AI jobs produces similar estimates (Appendix Table C.12). Using firms' global rather than U.S.-based AI-job shares also yields comparable results (Appendix Table C.13), addressing concerns that foreign AI employment is omitted or that changes in U.S. AI shares mechanically reflect relocation of non-AI jobs abroad. The absence of a relationship between AI investment and growth in U.S.-based Revelio employment count in Appendix Table C.6 provides additional evidence against an offshoring explanation.

Fourth, workers may retroactively add fashionable AI terminology to older job descriptions, particularly following ChatGPT's release (Bloom et al., 2026). This would introduce measurement error into our AI measure and attenuate our estimates. Several tests alleviate this concern. We find positive effects for Wave I AI jobs, which exclude generic terms like "AI" and predate the generative AI vocabulary; our measure is strongly correlated with the pre-2022 Cognism resume measure in Babina et al. (2024); and an earlier vintage of Revelio from October 2025 produces nearly identical estimates (Appendix Table C.14). A related concern is that since our skill list pools job postings over 2010–2025, recent AI vocabulary could be applied retroactively to earlier resumes; re-identifying AI jobs in each year using only that year's job postings data leaves the estimates unchanged (Appendix Table C.15).

Fifth, acquisitions could mechanically raise measured productivity if AI-investing firms disproportionately acquire asset-light or intangible-intensive targets. The results are unchanged when AI jobs at acquired subsidiaries are excluded (Appendix Table C.16), and AI investment is unrelated to acquired or total balance-sheet intangibles, which typically reflect externally purchased intangibles (Appendix Table C.17). This suggests that the productivity gains reflect organic improvements within firms rather than acquisition-driven compositional effects.

Sixth, while our baseline specification already excludes firms in the tech sector and semiconductor manufacturing, we further exclude the remaining Magnificent 7 firms (Apple and Tesla) and find similar results (Appendix Table C.18), indicating that our

results are not driven by a handful of large, technology-oriented firms.[25]

Finally, while the results in Section 3.2 suggest that reverse causality is unlikely, we test a broad set of contemporaneous confounders. (i) AI-investing firms may experience positive productivity shocks or already be on a higher growth trajectory. Appendix Table C.19 controls for initial levels of log sales per worker and revenue TFP, as well as state fixed effects to absorb state-level shocks, and finds similar results. We further confirm in Appendix Table C.20 that our results are robust to controlling for pre-trends in sales, employment, and TFP over 2010–2018, providing additional assurance that they are not driven by differential pre-existing trajectories among AI-investing firms. (ii) Appendix Table C.21 adds detailed industry fixed effects to absorb industry-level productivity shocks and hold fixed cross-industry differences in technology, input use, and market conditions. The results are nearly identical even with granular 5-digit NAICS fixed effects. (iii) Because our sample period spans the pandemic, we show in Appendix Table C.22 that our results are robust to controlling for firms' Covid exposure (Hassan et al., 2023) and for the share of remote-suitable jobs (Dingel and Neiman, 2020). (iv) AI investments may be correlated with investments in related technologies such as IT. We measure firms' IT investment using the share of jobs in computer- and IT-related occupations excluding AI jobs. Appendix Table C.23 shows that neither the initial level nor the change in non-AI IT investment is significantly associated with productivity growth over our sample period, and that controlling for these variables leaves the estimated effects of AI investment unchanged.[26] These additional tests address the most likely confounders, but they cannot rule out unobserved shocks entirely. We therefore turn to an instrumental variable strategy next.

3.4. Instrumental Variable Estimates

We instrument firm-level AI investment with each firm's ex-ante exposure to the supply of AI-trained graduates from universities with historically strong AI research following Babina et al. (2024). Because AI-skilled labor is a key constraint on firm investment in AI, firms with greater historical hiring ties to these universities have better access to AI talent and are better positioned to invest in AI. This instrument and its variants have been used in published (e.g., Krupa and Mullaney, 2026; McLemore and Mihov, 2026) and ongoing work (e.g., Hampole et al., 2025). Appendix D provides full construction and validation details.

[25] Five of the Magnificent 7 firms are already excluded from our baseline analysis: Microsoft, Alphabet, and Meta are in the tech sector; Nvidia is in the semiconductor manufacturing industry; and Amazon has over half of its jobs outside the U.S. in 2024.

[26] Because we find consistent results in manufacturing and non-manufacturing sectors, it is unlikely that our estimates are driven by robot adoption, which is concentrated in manufacturing.

We define AI-strong universities based on their pre-2010 share of AI researchers in the Open Academic Graph publications data, predating the commercial rise of AI. This definition comes from Babina et al. (2024), which provides a detailed description of the methodology. Using Revelio, we then construct each firm's historical hiring networks to these universities. Validating our classification, AI-strong universities subsequently produce substantially more AI-trained graduates (Appendix Figure D.2). The strategy requires firms to recruit from differentiated and stable sets of universities, and both conditions hold: firms tend to draw talent from a concentrated group of schools,[27] and their pre-2010 hiring networks strongly predict their post-2018 recruiting (Appendix Table D.3).

One concern is whether pre-2010 AI research is relevant to the technologies driving AI investment in our sample period, particularly generative AI. AI progress is highly cumulative: modern generative and agentic systems build on earlier advances in neural networks, statistical learning, computer vision, and NLP before 2010.[28] Consistent with this continuity, a substantial share of generative and agentic AI jobs mention ML, NLP, or computer vision keywords, indicating that AI human capital carries across technology waves. Our measure therefore captures durable institutional depth in AI research—the faculty, laboratories, and doctoral pipelines that continued to train workers throughout both waves, as illustrated in Appendix Figure D.2.

The instrument is the share of a firm's pre-2010 STEM hires drawn from AI-strong universities:

$$IV_i = \sum_u s_{iu}^{Pre2010} AIstrong_u,$$

where $s_{iu}^{Pre2010}$ is the share of STEM workers in firm i between 2005 and 2010 who had their most recent degree (undergraduate or graduate) from university u,[29] and $AIstrong_u$ equals one if university u is an AI-strong university based on pre-2010 publications.

The exclusion restriction requires that a firm's pre-2010 ties to AI-strong universities,

[27]The average firm sources 19% of its fresh graduates from a single dominant university, 47% from its top five universities, and 61% from its top ten. By comparison, the largest university accounts for only 1.6% of the fresh graduates nationwide, the five largest for 7.1%, and the ten largest for 12.9%.

[28]For example, Raina et al. (2009) showed that graphics processors could train deep models orders of magnitude faster than conventional hardware—the approach behind every large language model trained during our sample period—while the ImageNet database (Deng et al., 2009) established the benchmark that drove the computer vision advances now embedded in multimodal systems; both came from universities our measure classifies as AI-strong. The researchers who built the first wave also led the second: Fei-Fei Li, who led ImageNet, founded the generative AI firm World Labs in 2024; Ilya Sutskever, a co-author of AlexNet, went on to serve as chief scientist of OpenAI; and Yann LeCun, whose convolutional networks predate the deep learning era, was Meta's chief AI scientist from 2018 to 2025.

[29]STEM workers are workers with STEM majors in college, including biology, chemistry, engineering, information technology, mathematics, physics, and statistics.

measured well before the rise of commercial AI adoption, affect outcomes over 2018–2024 only through exposure to the supply of AI-trained labor. One potential threat is that these ties also proxy for access to technically skilled or elite graduates more broadly. We therefore control separately for firms' historical exposure to computer-science-strong and top-20 universities. A second concern is geographic: firms located in technology hubs may both hire from nearby AI-intensive universities and benefit from regional agglomeration effects. Our baseline sample already excludes tech-sector firms, which are most susceptible to this channel, and specifications with state fixed effects further absorb regional differences.

Appendix D provides additional support for the identifying assumption: firms do not increase their hiring from AI-strong universities during the pre-period in anticipation of future AI needs, and historic exposure to AI-strong universities does not predict higher productivity growth either before 2010 or over 2010–2018.

Appendix Table D.4 presents the first-stage results.[30] All specifications include the full set of baseline controls and exposure to CS-strong and top-20 universities. We then progressively introduce (i) state fixed effects, which absorb local labor market conditions and regional agglomeration effects, and (ii) pre-period growth in firm sales, employment, and TFP from 2010 to 2018, which helps account for unobserved firm-level attributes that could jointly influence productivity trajectories and the propensity to hire AI talent. The instrument has a strong first stage with F-statistics above 10 for all specifications, suggesting that firms' pre-2010 exposure to AI-strong universities strongly predicts their AI investment over 2018–2024.

Table 4 reports the 2SLS estimates. Across specifications, instrumented AI investment significantly raises both sales per worker and revenue TFP. With the full set of controls, a one-standard-deviation increase in AI investment raises sales per worker by 14.4% and revenue TFP by 11.4%.[31] These results support a causal effect of firms' internal AI investments on productivity.

4. Mechanisms

In this section, we examine the mechanisms linking firms' AI investments to productivity growth in recent years.

[30]The number of observations is smaller than in the OLS regressions because we only keep firms for which over half of the STEM workers between 2005 and 2010 have non-missing university information and had their most recent degree from one of the 659 universities with publications data.

[31]The IV estimates exceed the OLS estimates, although the difference is significant only at the 10% level for sales per worker and insignificant for TFP. This may reflect attenuation from measurement error in our AI measure or downward bias in OLS if firms with weaker expected productivity growth invest more in AI. Consistent with the latter, survey evidence from Acemoglu et al. (2023) shows that firms adopting advanced technologies tend to be larger and already in a lower-growth phase of their life cycle.

4.1. Firm AI Investment: Tangible versus Intangible Capital

Investments in new technologies can raise labor productivity by deepening tangible capital, by building intangible capital, or both. General-purpose technologies in particular often require not only investment in physical capital, but also complementary investments in organizational change, worker skills, and redesigned business processes (Bresnahan and Trajtenberg, 1995; Brynjolfsson et al., 2002; Bartel et al., 2007). AI may operate on both margins: computing and data storage equipment on the tangible side, but experimentation, data curation, training, workflow redesign, and integration on the intangible side (Corrado et al., 2009). Whether AI-investing firms tilt toward intangibles versus tangibles is therefore a first-order test of which channel dominates. We therefore re-estimate our main specification with tangible (PP&E) and intangible (SG&A, R&D) investment as outcomes over 2018–2024. Appendix Table C.24 shows that AI-investing firms invest more in intangible but not in tangible capital: SG&A rises by roughly 5% and R&D by 2.5%, while the PP&E coefficient is small and insignificant (-0.7%).

Because intangible capital is embodied in skilled labor, workforce composition provides a proxy for firms' stock of intangible knowledge and capabilities. Using the Revelio job-worker-firm-year panel, we construct firm-year employment counts and shares by job seniority and highest education level. Appendix Table C.25 repeats the main specification using changes in education and seniority shares as outcomes, excluding AI-related workers to avoid a mechanical relationship since AI-related jobs tend to be more senior and highly educated. AI-investing firms increase the shares of workers with advanced degrees (master's and PhDs) and in senior roles, while the shares of workers without college degrees or in junior roles decline.[32] These patterns confirm the conjecture in Aghion et al. (2019) and findings in Babina et al. (2025) that more AI-intensive firms employ a higher share of high-skill workers, and are consistent with recent evidence that AI adoption disproportionately displaces junior workers (e.g., Brynjolfsson et al., 2025a; Hosseini Maasoum and Lichtinger, 2025; Kim and Koning, 2026). Average wages also rise (Appendix Table C.27), indicating a shift toward high-skilled labor.

Together, the rise in SG&A, R&D, and high-skilled employment points to intangible-capital accumulation as the dominant margin among the non-tech AI-investing firms in our sample. This pattern contrasts with previous technological advances, which spurred tangible investment and capital deepening among adopters (Brynjolfsson and Hitt, 1996; Acemoglu et al., 2020; Benmelech and Zator, 2025). This does not imply that AI requires little physical capital. Large investments in data centers and specialized hardware are

[32] Appendix Table C.26 examines employment levels and shows that the total employment of highly educated and senior workers also significantly increases as firms invest in AI.

concentrated upstream among hyperscalers, while most non-technology firms rent this infrastructure rather than build it themselves.[33] Their AI investment is therefore more likely to manifest as intangible rather than tangible capital accumulation.

4.2. Firm AI Investment and Organization Capital

An important component of intangible capital is organization capital (OC) (Peters and Taylor, 2017). Given that firm AI investments are associated with a significant increase in SG&A—a key indicator of investment in OC—we next examine the role of OC in linking AI investments to productivity growth.

Organization capital is durable, firm-specific knowledge accumulated through learning-by-doing that allows firms to operate more efficiently (Prescott and Visscher, 1980). It is built by workers and embodied in organizational routines, systems, and key employees (Eisfeldt and Papanikolaou, 2013; Crouzet et al., 2022). Firms take a long time to build OC (Atkeson and Kehoe, 2005), but the faster a firm can learn, the quicker it accumulates OC and realizes productivity gains.

Several features distinguish OC from related intangible assets. OC is *built rather than purchased*, accumulating gradually through firms' learning from operations rather than being bought on the market. OC is *embedded rather than separable*, residing in platforms, systems, routines, workflows, and coordination capabilities, which distinguishes it from the inputs to intangible capital (e.g., R&D expense) and the outputs they generate (e.g., patents). OC is *durable rather than ephemeral*, representing reusable knowledge rather than one-off outputs. Finally, OC is *organization-specific rather than general*, which cuts two ways. Within the firm, the asset is non-rival (Mihet and Philippon, 2019; Jones and Tonetti, 2020): an established routine or system serves many employees, units, and tasks at once without depletion.[34] Outside the firm it travels poorly: routines tuned to one organization often lose value elsewhere. These features distinguish OC from related concepts such as R&D, IT, customer capital, co-invention, and product, process, or architectural innovation, while absorptive capacity and architectural competence are specific sub-types of OC. Appendix B.1 discusses these comparisons.

What is special about AI technologies is that they are fundamentally "learning machines" (Turing, 1950). Internal AI investment can accelerate a firm's learning from its

[33] Consistent with this view, the FTC identifies Alphabet, Amazon, and Microsoft as the largest cloud service providers involved in major generative-AI partnerships, and McKinsey estimates that hyperscalers will account for roughly 70 percent of forecast U.S. data-center capacity through owned or leased infrastructure (Federal Trade Commission, Office of Technology Staff, 2025; Arora et al., 2025).

[34] Knowledge-hierarchy models make this leverage explicit (Garicano, 2000): organizations economize on costly knowledge acquisition by concentrating expertise in a few workers and referring exceptions to them, at a communication cost. Codifying that expertise into a firm-wide system reduces the need for the matching that the hierarchy exists to perform.

operations—how to improve workflows, respond to customer needs, and redesign processes and products. Models trained on firm-specific data and embedded in firm-specific systems and workflows become part of OC because much of their value depends on the organization in which they are deployed. AI workers train and maintain these models in collaboration with employees who possess domain knowledge of the firm's operations. The resulting asset is therefore not just the model, but the durable routines and organizational knowledge built around it that make the firm more productive. This mechanism helps explain why AI-related jobs are nearly twice as likely as non-AI jobs to involve building and maintaining firm-specific organizational assets.

To formalize this mechanism and derive testable predictions, Appendix E develops a model extending the firm life-cycle framework of Atkeson and Kehoe (2005). The key state variable, *organization capital*, is the durable asset a firm accumulates by learning from its own operations—procedures, structured data, trained models, and codified workflows that the firm continues to use. Productivity is increasing in the stock of OC, so a firm with more of it produces more from the same inputs. Because OC is built from the firm's operations, the flow into the stock is proportional to their scale. AI enters as a *codification technology* that raises the rate at which a firm's learning from its operations is converted into OC. For example, operational records—service transcripts, code repositories, and clinical documentation—can be transformed into trained models and systems that make firm-specific knowledge available across the organization (Singhal et al., 2023; Brynjolfsson et al., 2025b). AI therefore allows knowledge that would otherwise diffuse slowly through experience, mentoring, or gradual documentation to be codified and replicated more broadly within the firm.

The model yields two testable predictions. First, AI's productivity effect should emerge slowly even without adjustment costs or implementation lags. Because OC is a stock and AI raises only the flow into it, OC's accumulation remains gradual even when AI accelerates its buildup. Second, productivity gains should be largest for firms that start with relatively little OC for their size, as they have more scope for AI-driven OC accumulation. Think of OC as a firm's operating playbook: firms with a thin playbook have more processes, routines, and knowledge left to codify and improve, whereas firms with an already extensive playbook have less left to add.

To illustrate how AI accelerates the accumulation of OC, Appendix B.3 presents case studies of eight firms across sectors and seventeen of the OC-building job categories in Table 1. How AI performs the conversion is largely common across them: a model is trained on records firms already produce—research documents, meeting transcripts, sensor readings, or project records—turning documentation of past activity into an asset that can be queried, simulated against, or acted on. What AI converts differs sharply. At

Morgan Stanley, it is professional judgment: an assistant built on the firm's own research library, and a tool that takes notes during advisor meetings and writes them into the CRM, capturing a record that previously left no trace. At Procter & Gamble, it is physical process knowledge: plant telemetry becomes a digital twin against which process changes are trialed without halting the line. At Home Depot, it is the expertise of store associates—knowledge that had been available only in one aisle of one store at a time—assembled into a knowledge base available everywhere at once. In each case, the raw material is the firm's own operational record, the conversion is performed by the firm's own AI workers—precisely the roles our measure captures—and the resulting asset is useful to many workers inside the firm, while typically being worth considerably less outside it.

Next, we examine whether the data are consistent with the model's predictions. The first prediction matches the timing documented in Section 3.2: AI's productivity effects do not appear all at once but build gradually. Because OC is a stock, a firm that raises its rate of learning from operations does not see the gain immediately: under plausible depreciation rates, the model implies that productivity gains from AI investment can take five years or longer to reach half their eventual size. Two features of our setting reinforce this pattern. Most firms began investing intensively in AI only after 2015, so much of the early period contains little treatment to respond to; and the composition of AI jobs has shifted over time toward roles that are more likely to build OC (Figure 6). Both help explain why we find significant productivity growth among AI-investing firms in the later, but not the earlier, part of our sample.

Second, Table 5 examines whether firms' AI investments are associated with growth in the standard capitalized-SG&A measure of OC (Eisfeldt and Papanikolaou, 2013). As with productivity, there is no significant OC buildup among AI-investing firms over 2010–2018, but a significant increase over 2018–2024. Lead-lag tests show that past AI investment predicts subsequent OC growth, whereas past OC growth does not predict subsequent AI investment (Appendix Tables C.28 and C.29). The direction of predictability thus runs from AI investment to OC, mirroring what we find for productivity, and confirming the predictions of our model: firm productivity growth becomes visible when AI investments help increase the stock of OC substantially. Hence, these results suggest a mechanism for the long lag between AI investment and measurable productivity growth: the intermediate step of accumulating OC is itself slow.[35]

Third, we split AI jobs into OC-building and non-OC-building components. If AI investment raises productivity by accelerating OC accumulation, productivity gains should

[35] In Appendix Table C.30, the IV estimates support a causal interpretation of the effect of AI investment on OC growth.

be concentrated in AI jobs that build OC rather than those that do not. This is a sharper test than the aggregate relationship between firms' overall AI investment and productivity, since it compares alternative uses of AI investment rather than differences in its level. Table 6 strongly supports this prediction: growth in the share of OC-building AI jobs is positively associated with productivity growth, while growth in non-OC-building AI employment is not (this null result also holds when we omit OC-building AI jobs from the specification). Including both components jointly leaves the OC-building coefficient essentially unchanged. The two coefficients differ significantly at the 5% level, and each component accounts for roughly half of all AI jobs under the baseline threshold, so the null result on non-OC-building jobs is unlikely to reflect noise.[36] The results are also not driven by OC-building jobs being higher-quality: as shown in Section 2.2 and Appendix B.2, OC-building jobs do not have higher-quality job descriptions, nor are they more likely to be core AI development roles; what distinguishes them is concrete evidence of building systems the firm continues to use.

Consistent with this interpretation, OC-building AI jobs also predict faster growth in OC, proxied by capitalized SG&A expenses, while non-OC-building AI jobs do not (Appendix Table C.32). Thus, the same OC-building AI jobs that predict productivity growth also predict growth in the stock of OC through which the mechanism operates. One concern is that firms already rich in OC may both hire more OC-building AI workers and experience faster productivity growth for reasons unrelated to AI. Our estimates are virtually unchanged, however, when we control for firms' initial level of OC (Appendix Tables C.33 and C.34).

Finally, the model's second prediction is that firms with a low ex-ante stock of OC relative to their size should experience faster growth in both OC and productivity following AI investments. We test this by splitting firms at the median of pre-period OC scaled by PP&E and interacting AI investment with the two groups.[37] Table 7 reports productivity growth and Appendix Table C.35 reports growth in OC. The pattern matches the prediction on both margins. For productivity, initially low-OC firms show coefficients of 0.092 for sales per worker and 0.105 for TFP, both significant, while the corresponding high-OC coefficients are about half as large and insignificant. For OC, the coefficient for low-OC firms is roughly two to three times larger than for high-OC firms. Relatedly, Crouzet et al. (2026) show that old firms lag young ones when new technologies require a shift in organizational workstyles. In their account, an old firm's existing organization conflicts

[36] Appendix Table C.31 finds similar results with alternative thresholds of 0.5 and 0.7 for OC-building jobs.

[37] Because the model's prediction concerns OC relative to size, and because size itself is one of the outcomes, we include size-tercile fixed effects (interacted with industry fixed effects and an above-median-OC indicator) so the comparison holds scale fixed.

with the new technology; in ours, an established stock of OC simply leaves less to codify: OC-rich incumbents still invest and build, but have less scope for further accumulation.

A natural concern is mean reversion: firms classified as low-OC because of a temporary negative shock may subsequently grow faster mechanically. Two features of our setting limit this. First, identification comes from the interaction with AI investment rather than from group means, so generic reversion is absorbed by the group main effect. Second, OC is an accumulated stock that depreciates slowly, so any single year's flow shock moves it only marginally. We nonetheless address the concern directly. We first sort on OC/PP&E averaged over 2014–2018 rather than a single year, and the results are unchanged (Appendix Table C.36). We also estimate the interacted specification on pre-period growth over 2012–2018 in Appendix Table C.37. Mechanical reversion implies the pre-period interaction carries the opposite sign; a pre-trend or anticipatory response would instead produce the same sign. We find neither: the pre-period coefficients are small and insignificant for both groups. The heterogeneity therefore reflects differences in adjustment speed rather than mean reversion.

Taken together, the evidence points to OC as a central channel linking AI investment to productivity. AI-investing firms accumulate OC faster, and past AI investment predicts future OC growth while the reverse does not hold. The productivity gains arrive gradually, are concentrated in AI jobs that build OC rather than in other AI jobs, and are largest at firms with the least OC to begin with. These patterns suggest that AI raises productivity by generating firm-specific knowledge—embodied in systems, routines, and firm-trained models—that enables firms to learn faster and operate more efficiently.

Our findings speak to the literature on long productivity lags from general-purpose technologies (David, 1990; Bresnahan and Trajtenberg, 1995). There, the technology alone is insufficient: gains appear only after firms undertake complementary co-investments in reorganization, process redesign, and new skills (Bresnahan et al., 2002; Brynjolfsson and Hitt, 2003; Bloom et al., 2012), which are intangible and largely unmeasured (Brynjolfsson et al., 2021). An important part of these investments can be understood as OC—the routines, systems, and firm-specific knowledge that must be accumulated gradually to make the new technology productive. Our results sit squarely in this tradition: AI's productivity gains are mediated by OC and take years to appear. What we add is a second role for OC: because AI is a learning technology, it is not only an input that requires complementary OC but also a technology for producing OC, raising the rate at which a firm accumulates organizational knowledge from operations. This mechanism provides a further reason why productivity gains can be slow even when organizational adjustment itself is not costly: because OC is learned from experience, its accumulation takes time.

4.3. Alternative Mechanism: Task Automation and Labor Replacement

A common argument holds that AI can substitute for certain worker tasks, potentially displacing some jobs and increasing overall productivity (Acemoglu et al., 2022; Hampole et al., 2025). If this channel were driving our findings, we would expect roughly flat sales alongside falling employment. Neither pattern appears in our estimates—sales rise while employment does not decline. Moreover, under this mechanism, we would expect AI-investing firms to shed the workers most exposed to AI—who, according to standard exposure measures, tend to be highly skilled, technical, educated, and higher paid. Our results point in the opposite direction: against a flat total headcount, AI investments are associated with an increase in both the employment levels and shares of highly educated workers (Appendix Tables C.25 and C.26) as well as average wages (Appendix Table C.27), and in unreported results, we find that the share of STEM workers also increases in AI-investing firms.

These findings do not rule out labor-replacing effects of AI in other settings; instead, they imply that this mechanism does not drive productivity effects in our setting. Two features of our setting explain why. First, our measure captures the net effect of AI across all applications, not only the labor-replacing ones; second, we focus on internal firm AI development and implementation, whereas AI labor replacement effects may be more important for outsourced AI tools like chatbots. Consistent with this distinction, large-scale usage evidence suggests that workplace AI use remains shallow and predominantly collaborative: AI is used for roughly a fifth of tasks in the median occupation with any AI use, and end-to-end task automation is limited in scope (Iscenko et al., 2026).

A related possibility is that AI—and generative AI in particular—replaces some tasks within jobs and reallocates workers toward other activities, thereby raising their productivity (e.g., Noy and Zhang, 2023). This also appears unlikely to explain our results. We find that the productivity gains are not confined to generative and agentic AI: Wave I technologies—machine learning, NLP, and computer vision, which predate the generative-AI tools through which this channel would operate—are independently associated with productivity growth. The timing also cuts against this channel: task-level reallocation should raise output on impact, whereas the responses we estimate build over several years, which is what a stock-flow mechanism delivers and within-job task substitution does not. Moreover, only around 1% of AI-related jobs in our sample are non-AI roles using AI tools as productivity aids, whereas the majority involve building or implementing internal AI capabilities. The timing and composition of AI investment therefore fit gradual OC accumulation better than immediate within-job task substitution.[38]

[38] Demirer et al. (2026b) find that large task-level gains from AI coding tools attenuate substantially before

4.4. Alternative Mechanism: Firm AI Investment and Price Increases

Because our productivity measures are revenue-based, higher measured productivity could reflect price increases rather than real output growth. A related possibility is that AI-investing firms capture market share from rivals: business stealing raises revenue at adopting firms even absent any price change, and likewise need not reflect an outward shift of the productive frontier.

Lacking firm-level data on prices and quantities, we use industry-level data to assess both channels, which have distinct implications for industry prices and productivity. If AI-investing firms simply charge more, industry-level prices should rise, and by enough to account for the measured revenue productivity gains. If their gains instead reflect transfers from competitors, they should not aggregate into industry-level productivity improvements. Gains from adopting firms producing more, by contrast, should survive aggregation.

We estimate analogous long-differences regressions at the 5-digit NAICS industry level, relating 2018–2024 changes in industry productivity or producer price indices (PPI) to contemporaneous changes in the industry-level AI worker share, with sector fixed effects and weighting by 2018 industry resume counts. Aggregating to the industry level allows us to include firms that entered after 2018 or exited before 2024, which are excluded from the firm-level panel.

Appendix Table C.38 shows that industries with higher AI investments experience significantly higher productivity growth over 2018–2024, mirroring the firm-level results.[39] The last two columns show that industries that invest heavily in AI exhibit no detectable change in price indices, and the confidence interval rules out price increases large enough to account for the productivity gains in columns 1–4. Because the PPI is an industry-level index, it would mask relative price movements if AI-investing firms raised prices while their competitors lowered them; the aggregation results in columns 1–4 weigh against that case, since such a reshuffling of revenue would leave industry productivity unchanged.[40]

In sum, the industry-level evidence is inconsistent with the price explanation, and points instead to AI investments expanding real output.

---

reaching completed projects and releases, and Demirer et al. (2026a) show theoretically that such gains pass through only once firms re-bundle tasks into jobs. Both imply that task-level productivity is a poor guide to firm productivity absent organizational adjustment.

[39] We find no productivity effects at the industry level in the earlier period from 2010 to 2018.

[40] The industry AI share in Appendix Table C.38 is constructed from publicly listed firms alone. In Appendix Table C.39, we reconstruct the AI share using all public and private firms in the Revelio data and confirm that price indices move little with this broader measure of AI investment. We also find that industry-level total employment follows a similar pattern to the firm-level results.

5. Conclusion

Whether AI raises productivity is the central question about its economic consequences, and the growth literature entertains a wide range of outcomes (Trammell and Korinek, 2026). Using new measures of firms' internal AI investments through 2024—covering both the first wave of machine learning, natural language processing, and computer vision and the recent wave of generative and agentic AI—we show that U.S. public firms investing more in AI experience significantly faster growth in productivity between 2018 and 2024, in sharp contrast to the absence of such gains over 2010–2018.

The evidence points to organization capital being a central mechanism. AI-investing firms accumulate more OC as measured by capitalized SG&A. The productivity gains are concentrated in AI jobs that build organization capital—firm-specific knowledge, systems, and processes—and absent in AI jobs that do not. Both productivity and OC growth are also largest among firms that begin with relatively little OC, as our model predicts. Whereas the general-purpose-technology literature emphasizes OC as a complement firms must build around new technologies, our results suggest that AI technologies can also produce OC by accelerating the conversion of operational learning into durable organizational assets.

Our findings suggest cautious optimism. Our evidence of increased productivity from AI comes from U.S. public firms, which have been investing in AI at roughly twice the rate of other U.S. employers. These firms at the frontier of AI investment serve as canaries—not in a coal mine warning of danger, but in a gold mine offering an early signal of AI's broader potential and of the organization capital that firms must build to realize it. If the history of general-purpose technologies is a guide, these gains may spread as more firms invest in AI and use it to build the organization capital that raises productivity.

References

Abis, S., Veldkamp, L., 2024. The Changing Economics of Knowledge Production. Review of Financial Studies 37, 89–118.

Abraham, K.G., Hershbein, B., Houseman, S.N., Truesdale, B.C., 2024. The Independent Contractor Workforce: New Evidence on Its Size and Composition and Ways to Improve Its Measurement in Household Surveys. ILR Review 77, 336–365.

Abraham, K.G., Houseman, S.N., Truesdale, B.C., 2026a. Are the Self-Employed Really Their Own Boss?: Evidence on the Degree of Autonomy and Control the Self-Employed Have Over Their Work. Upjohn Institute Policy Paper 2026-038.

Abraham, K.G., Houseman, S.N., Truesdale, B.C., 2026b. The "Missing Self-Employed": How Better Data Change the Picture of Self-Employment in the United States. Upjohn Institute Policy Paper 2026-039.

Acemoglu, D., Anderson, G.W., Beede, D.N., Buffington, C., Childress, E.E., Dinlersoz, E., Foster, L.S., Goldschlag, N., Haltiwanger, J.C., Kroff, Z., Restrepo, P., Zolas, N., 2023. Advanced Technology Adoption: Selection or Causal Effects? AEA Papers and Proceedings 113, 210–214.

Acemoglu, D., Autor, D., Hazell, J., Restrepo, P., 2022. Artificial Intelligence and Jobs: Evidence from Online Vacancies. Journal of Labor Economics 40, S293–S340.

Acemoglu, D., Lelarge, C., Restrepo, P., 2020. Competing with Robots: Firm-Level Evidence from France. AEA Papers and Proceedings 110, 383–388.

Acemoglu, D., Restrepo, P., 2020. Robots and Jobs: Evidence from U.S. Labor Markets. Journal of Political Economy 128, 2188–2244.

Acharya, V.V., Shao, X., Xu, Z., 2026. AI and Productivity: The Role of Innovation. Working Paper.

Acikalin, U.U., Caskurlu, T., Hoberg, G., Phillips, G.M., 2026. Intellectual Property Protection Lost and Competition: An Examination Using Large Language Models. Journal of Financial Economics 182.

Aghion, P., Howitt, P., 1992. A Model of Growth Through Creative Destruction. Econometrica 60, 323–351.

Aghion, P., Jones, B.F., Jones, C.I., 2019. Artificial Intelligence and Economic Growth, in: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 237–282.

Agrawal, A., Gans, J., Goldfarb, A., 2022. Prediction Machines, Updated and Expanded: The Simple Economics of Artificial Intelligence. Harvard Business Review Press.

Akcigit, U., Chikis, C.A., Dinlersoz, E., Goldschlag, N., 2026. Attention (and Money) Is All You Need: Why Universities Are Struggling to Keep AI Talent. NBER Working Paper 34964.

Aldasoro, I., Gambacorta, L., Pál, R., Revoltella, D., Weiss, C., Wolski, M., 2026. AI Adoption, Productivity and Employment: Evidence from European Firms. BIS Working Paper 1325.

Argote, L., Beckman, S.L., Epple, D., 1990. The Persistence and Transfer of Learning in Industrial Settings. Management Science 36, 140–154.

Arora, C., Sorel, M., Sachdeva, P., Bhan, A., He, J., Shaw, N., Garg, R., Ravishankar, S., 2025. The Next Big Shifts in AI Workloads and Hyperscaler Strategies. McKinsey & Company. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/the-next-big-shifts-in-ai-workloads-and-hyperscaler-strategies.

Atkeson, A., Kehoe, P.J., 2005. Modeling and Measuring Organization Capital. Journal of Political Economy 113, 1026–1053.

Autor, D., 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. Journal of Economic Perspectives 29, 3–30.

Autor, D., Dorn, D., Hanson, G.H., 2013. The China Syndrome: Local Labor Market Effects of Import Competition in the United States. American Economic Review 103, 2121–68.

Babina, T., 2026. Understanding Firms' AI Efforts and Their Economic Impact. The Review of Corporate Finance Studies, cfag027.

Babina, T., Fedyk, A., He, A., Hodson, J., 2024. Artificial Intelligence, Firm Growth, and Product Innovation. Journal of Financial Economics 151, 103745.

Babina, T., Fedyk, A., He, A., Hodson, J., 2025. Firm Investments in Artificial Intelligence Technologies and Changes in Workforce Composition, in: Technology, Productivity, and Economic Growth. University of Chicago Press, pp. 75–118.

Baily, M.N., Byrne, D.M., Kane, A.T., Soto, P.E., 2025. Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope? Finance and Economics Discussion Series 2025-053. Board of Governors of the Federal Reserve System.

Barbaschow, A., 2021. Moderna Leveraging Its "AI Factory" to Revolutionise the Way Diseases Are Treated. ZDNet. https://www.zdnet.com/article/moderna-leveraging-its-ai-factory-to-revolutionise-the-way-diseases-are-treated/.

Barrios, J.M., Campbell, J.L., Johnson, R.G., Liu, C., 2025. Artificially Intelligent or Artificially Inflated? Determinants and Informativeness of Corporate AI Disclosures. Working Paper.

Bartel, A., Ichniowski, C., Shaw, K., 2007. How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. Quarterly Journal of Economics 122, 1721–1758.

Baslandze, S., Edwards, Z., Graham, J.R., McClure, T., Sparks, M., Meyer, B., Waddell, S.R., Weitz, D., 2026. Artificial Intelligence, Productivity, and the Workforce: Evidence from Corporate Executives. Journal of Finance: Insights and Perspectives Forthcoming.

Baviskar, S., Branstetter, L., Drayton, C., Fujiy, B.C., Hovy, E., Tambe, P., Wu, L., 2026. Elite Talent, Invention, and Firm Productivity in the Age of AI. Working Paper.

Benkard, C.L., 2000. Learning and Forgetting: The Dynamics of Aircraft Production. American Economic Review 90, 1034–1054.

Benmelech, E., Zator, M., 2025. Robots and Firm Investment. Journal of Financial Economics 174, 104183.

Benton Institute for Broadband and Society, 2026. Verizon Taps AI Agents for Network Automation. June 12. https://www.benton.org/headlines/verizon-taps-ai-agents-network-automation.

Beraja, M., Talamàs, E., 2026. The Value of Organizational Learning Technologies. Working Paper.

Bick, A., Blandin, A., Deming, D.J., Fuchs-Schündeln, N., Jessen, J., 2026. Mind the Gap: AI Adoption in Europe and the US. NBER Working Paper 34995.

Blank, M., Schubert, G., Zhang, B., 2025. The Household Impact of Generative AI: Evidence from Internet Browsing Behavior. Working Paper.

Bloom, N., Moore, G., Simon, L.K., Wilkie-Rogers, C., 2026. Time Travel on Professional Profiles.

NBER Working Paper 35546.

Bloom, N., Sadun, R., Van Reenen, J., 2012. Americans Do IT Better: US Multinationals and the Productivity Miracle. American Economic Review 102, 167–201.

Blumenfeld, A., Hazell, J., Lian, C., Schaab, A., 2026. The Macroeconomic Effect of AI: Sizing the Software Engineering Channel. Working Paper.

Bonelli, M., 2026. Data-Driven Investors. Review of Financial Studies 39 (7), 1909–1969.

Bonney, K., Breaux, C.L., Dinlersoz, E., Foster, L.S., Haltiwanger, J.C., Pande, A.A., 2026. The Microstructure of AI Diffusion: Evidence from Firms, Business Functions, and Worker Tasks. NBER Working Paper 35141.

Breitung, C., Müller, S., 2025. Global Business Networks. Journal of Financial Economics 166, 104007.

Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M., 2002. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. Quarterly Journal of Economics 117, 339–376.

Bresnahan, T.F., Greenstein, S., 1996. Technical Progress and Co-invention in Computing and in the Use of Computers. Brookings Papers on Economic Activity. Microeconomics, 1–77.

Bresnahan, T.F., Trajtenberg, M., 1995. General Purpose Technologies: 'Engines of Growth'? Journal of Econometrics 65, 83–108.

Bryan, K.A., 2026. The Economic Impacts of Artificial Intelligence: A Multidisciplinary, Multi-book Review. Journal of Economic Literature 64, 281–300.

Brynjolfsson, E., Chandar, B., Chen, R., 2025a. Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence. Working Paper.

Brynjolfsson, E., Hitt, L.M., 1996. Paradox Lost? Firm-Level Evidence on the Returns to Information Systems Spending. Management Science 42, 541–558.

Brynjolfsson, E., Hitt, L.M., 2000. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. Journal of Economic Perspectives 14, 23–48.

Brynjolfsson, E., Hitt, L.M., 2003. Computing Productivity: Firm-Level Evidence. The Review of Economics and Statistics 85, 793–808.

Brynjolfsson, E., Hitt, L.M., Yang, S., 2002. Intangible Assets: Computers and Organizational Capital. Brookings Papers on Economic Activity 2002, 137–199.

Brynjolfsson, E., Li, D., Raymond, L.R., 2025b. Generative AI at Work. Quarterly Journal of Economics 140, 889–942.

Brynjolfsson, E., Rock, D., Syverson, C., 2019. Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics, in: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 23–57.

Brynjolfsson, E., Rock, D., Syverson, C., 2021. The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. American Economic Journal: Macroeconomics 13, 333–372.

Cai, K., 2025. Verizon Says Google AI for Customer Service Agents Has Led to Sales Jump. U.S. News, April 9. https://money.usnews.com/investing/news/articles/2025-04-09/verizon-says-google-ai-for-customer-service-agents-has-led-to-sales-jump.

Cai, W., Prat, A., Yu, J., 2026. Measuring Organizational Capital. Journal of Accounting Research.

Campello, M., Kankanhalli, G., Muthukrishnan, P., 2024. Corporate Hiring Under COVID-19: Financial Constraints and the Nature of New Jobs. Journal of Financial and Quantitative Analysis 59, 1541–1585.

Cao, S., Jiang, W., Wang, J., Yang, B., 2024. From Man vs. Machine to Man + Machine: The Art and AI of Stock Analyses. Journal of Financial Economics 160, 103910.

Cao, Y., Liu, M., Qiu, J., Zhao, R., 2025. Determinants and Information Content of Voluntary Artificial Intelligence Disclosure. Working Paper.

Challame, B., 2023. Moderna, Powered by AI: Our Journey to Becoming a Real-Time AI Organization. Moderna Blogs, November 29. https://www.modernatx.com/media-center/all-media/blogs/moderna-powered-AI.

Charoenwong, B., Kwan, A., Traina, J., 2026. The Composition of Intangible Capital: Evidence from LLM Classification of Corporate Disclosures. Working Paper.

Chatterji, A., Rakholia, N., Weeratunga, G., Holtz, D., 2026. How Organizations Use AI: Evidence from ChatGPT. Working Paper.

Chawla, M., 2025. The Real Effects of Dissemination of Corporate Disclosures on Innovation. Working Paper.

Chen, S., Sialm, C., Xu, D.X., 2026. The Growth and Performance of Artificial Intelligence in Asset Management. NBER Working Paper 35273.

Chequer, M., Herkenhoff, K., Papanikolaou, D., Rothbaum, J., Schmidt, L., Seegmiller, B., 2026. Labor, Despite AI. Working Paper.

Cockburn, I.M., Henderson, R., Stern, S., 2019. The Impact of Artificial Intelligence on Innovation: An Exploratory Analysis, in: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), The Economics of Artificial Intelligence: An Agenda. University of Chicago Press, pp. 115–146.

Cockburn, I.M., Henderson, R.M., 1998. Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery. The Journal of Industrial Economics 46, 157–182.

Cohen, W.M., Klepper, S., 1996. Firm Size and the Nature of Innovation within Industries: The Case of Process and Product R&D. The Review of Economics and Statistics 78, 232–243.

Cohen, W.M., Levinthal, D.A., 1989. Innovation and Learning: The Two Faces of R&D. The Economic Journal 99, 569–596.

Cohen, W.M., Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. Administrative Science Quarterly 35, 128–152.

Corrado, C., Hulten, C., Sichel, D., 2009. Intangible Capital and U.S. Economic Growth. Review of Income and Wealth 55, 661–685.

Crouzet, N., Eberly, J.C., 2019. Understanding Weak Capital Investment: The Role of Market Concentration and Intangibles. NBER Working Paper 25869.

Crouzet, N., Eberly, J.C., 2021. Intangibles, Markups, and the Measurement of Productivity Growth. Journal of Monetary Economics 124, S92–S109.

Crouzet, N., Eberly, J.C., 2023. Rents and Intangible Capital: A Q+ Framework. Journal of Finance 78, 1873–1916.

Crouzet, N., Eberly, J.C., Eisfeldt, A.L., Papanikolaou, D., 2022. The Economics of Intangible Capital. Journal of Economic Perspectives 36, 29–52.

Crouzet, N., He, Z., Lyonnet, V., Ma, Y., 2026. Why Don't Old Firms Do New Things? Working

Paper.

Cullen, Z., Li, D., Li, S., 2026. Labor as Capital: AI and the Ownership of Expertise. Working Paper.

CVS Health, 2026a. Aetna Reduces Claims Processing Time by More Than 20% with AI to Improve Care Experience. Press release, May 26. https://www.cvshealth.com/news/innovation/aetna-reduces-claims-processing-time-by-more-than-20-percent-with-ai-to-improve-care-experience.html.

CVS Health, 2026b. First Quarter 2026 Results. Form 8-K, Exhibit 99.1, U.S. Securities and Exchange Commission. https://www.sec.gov/Archives/edgar/data/0000064803/000006480326000051/cvs_ex99x1q1-26.htm.

Czarnitzki, D., Fernández, G.P., Rammer, C., 2023. Artificial Intelligence and Firm-Level Productivity. Journal of Economic Behavior & Organization 211, 188–205.

Davenport, T.H., Bean, R., 2025. How Procter & Gamble Uses AI to Unlock New Insights from Data. MIT Sloan Management Review. https://sloanreview.mit.edu/article/how-procter-gamble-uses-ai-to-unlock-new-insights-from-data/.

David, P.A., 1990. The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox. American Economic Review 80, 355–361.

Davis, S.J., Haltiwanger, J., Jarmin, R., Miranda, J., Foote, C., Nagypal, E., 2006. Volatility and Dispersion in Business Growth Rates: Publicly Traded Versus Privately Held Firms. NBER Macroeconomics Annual 21, 107–179.

Davis, S.J., Haltiwanger, J.C., Schuh, S., 1996. Job Creation and Destruction. MIT Press, Cambridge, MA.

De Loecker, J., Eeckhout, J., Unger, G., 2020. The Rise of Market Power and the Macroeconomic Implications. Quarterly Journal of Economics 135, 561–644.

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz, H., Kellogg, K.C., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K.R., 2026. Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. Organization Science 37, 403–423.

Deming, D., Kahn, L.B., 2018. Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. Journal of Labor Economics 36, S337–S369.

Demirer, M., Horton, J.J., Immorlica, N., Lucier, B., Shahidi, P., 2026a. Chaining Tasks, Redefining Work: A Theory of AI Automation. NBER Working Paper 34859.

Demirer, M., Musolff, L., Yang, L., 2026b. Writing Code vs. Shipping Code: Productivity Effects Across Generations of AI Coding Tools. NBER Working Paper 35275.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. ImageNet: A Large-Scale Hierarchical Image Database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE. pp. 248–255.

Dessein, W., Prat, A., 2022. Organizational Capital, Corporate Leadership, and Firm Dynamics. Journal of Political Economy 130, 1477–1539.

Dingel, J.I., Neiman, B., 2020. How Many Jobs Can Be Done at Home? Journal of Public Economics 189, 104235.

Eisfeldt, A.L., Hartman-Glaser, B., Kim, E.T., Lee, K.B., 2026a. Intangible Intensity. NBER Working Paper 34882.

Eisfeldt, A.L., Papanikolaou, D., 2013. Organization Capital and the Cross-Section of Expected Returns. Journal of Finance 68, 1365–1406.

Eisfeldt, A.L., Schubert, G., Zhang, M.B., Taska, B., 2026b. Generative AI and Firm Values. Journal of Finance Forthcoming.

Eloundou, T., Manning, S., Mishkin, P., Rock, D., 2024. GPTs are GPTs: Labor market impact potential of LLMs. Science 384, 1306–1308.

Fahlenbrach, R., Sanz, L., Stulz, R.M., 2026. Organization Capital, Large Startups, and the Dearth of IPOs. ECGI Finance Working Paper 1140/2026.

Falato, A., Kadyrzhanova, D., Sim, J.W., Steri, R., 2022. Rising Intangible Capital, Shrinking Debt Capacity, and the U.S. Corporate Savings Glut. Journal of Finance 77, 2799–2852.

Federal Trade Commission, Office of Technology Staff, 2025. Partnerships Between Cloud Service Providers and AI Developers: FTC Staff Report on AI Partnerships & Investments 6(b) Study. Technical Report. Federal Trade Commission.

Felten, E.W., Raj, M., Seamans, R., 2018. A method to link advances in artificial intelligence to occupational abilities. AEA Papers and Proceedings 108, 1–4.

Fortune, 2026. How Expedia's CTO Is Using AI to Transform Work for 17,000 Employees—and Travel for Millions. Fortune, January 14. https://www.fortune.com/2026/01/14/how-expedias-cto-is-using-ai-to-transform-work-for-17000-employees-and-travel-for-millions.

Foster, L., Haltiwanger, J., Syverson, C., 2016. The Slow Growth of New Plants: Learning about Demand? Economica 83, 91–129.

Garicano, L., 2000. Hierarchies and the Organization of Knowledge in Production. Journal of Political Economy 108, 874–904.

Gentzkow, M., Kelly, B., Taddy, M., 2019. Text as Data. Journal of Economic Literature 57, 535–574.

Gibbons, R., Henderson, R., 2012. Relational Contracts and Organizational Capabilities. Organization Science 23, 1350–1364.

Gofman, M., Jin, Z., 2024. Artificial intelligence, Education, and Entrepreneurship. Journal of Finance 79, 631–667.

Goldfarb, A., Taska, B., Teodoridis, F., 2023. Could Machine Learning be a General Purpose Technology? A Comparison of Emerging Technologies Using Data from Online Job Postings. Research Policy 52, 104653.

Grant, R.M., 1996. Toward a Knowledge-Based Theory of the Firm. Strategic Management Journal 17, 109–122.

Guardian, 2018. Revealed: Google's 'Two-tier' Workforce Training Document. https://www.theguardian.com/technology/2018/dec/11/google-tvc-full-time-employees-training-document.

Hampole, M., Papanikolaou, D., Schmidt, L.D.W., Seegmiller, B., 2025. Artificial Intelligence and the Labor Market. NBER Working Paper 33509.

Hassan, T.A., Hollander, S., van Lent, L., Schwedeler, M., Tahoun, A., 2023. Firm-Level Exposure to Epidemic Diseases: COVID-19, SARS, and H1N1. Review of Financial Studies 36, 4919–4964.

He, B., Mostrom, L.I., Sufi, A., 2024. Investing in Customer Capital. NBER Working Paper 33171.

Henderson, R.M., Clark, K.B., 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. Administrative Science Quarterly 35,

9–30.

Henderson, R.M., Cockburn, I.M., 1994. Measuring Competence? Exploring Firm Effects in Pharmaceutical Research. Strategic Management Journal 15, 63–84.

Hershbein, B., Kahn, L.B., 2018. Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. American Economic Review 108, 1737–1772.

High, P., 2026. How Expedia Is Reinventing Travel Through AI and Agentic Design. Forbes, May 29. https://www.forbes.com/sites/peterhigh/2026/05/29/how-expedia-is-reinventing-travel-through-ai-and-agentic-design/.

Hoberg, G., Phillips, G., 2016. Text-Based Network Industries and Endogenous Product Differentiation. Journal of Political Economy 124, 1423–1465.

Hosseini Maasoum, S.M., Lichtinger, G., 2025. Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Resume and Job Posting Data. Working Paper.

Humlum, A., Vestergaard, E., 2026. Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI. NBER Working Paper 33777.

Imas, A., 2026. What is the Impact of AI on Productivity? https://aleximas.substack.com/p/what-is-the-impact-of-ai-on-productivity. Substack post.

Imrohoroğlu, A., Tüzel, Ş., 2014. Firm-Level Productivity, Risk, and Return. Management Science 60, 2073–2090.

Iscenko, Z., Strand, S., Chen, Y., Aimard, G., Codreanu, M., Sampathkumar, V., Imas, A., Jacobs, J., Muiruri, E., Mateos-Garcia, J., Ng, J.J., Javed, S., Martin, J., Ajmeri, O., Calin, D., Kim, A., Curto Millet, F., Manyika, J., 2026. Google's AI & Economy ATLAS v1.0: Mapping Gemini Usage in the Economy. Working Paper.

Jiang, W., Park, J., Xiao, R.J., Zhang, S., 2025. AI and the Extended Workday: Productivity, Contracting Efficiency, and Distribution of Rents. NBER Working Paper 33536.

Jones, C.I., Tonetti, C., 2020. Nonrivalry and the Economics of Data. American Economic Review 110, 2819–58.

Kalyani, A., Bloom, N., Carvalho, M., Hassan, T.A., Lerner, J., Tahoun, A., 2025. The Diffusion of New Technologies. Quarterly Journal of Economics 140, 1299–1365.

Kim, H., Koning, R., 2026. AI-Native Firms. Working Paper.

Kitishian, D., 2025. Nike's AI Strategy: Analysis of Dominance in Athletic Apparel, Footwear. Klover.AI research report. https://www.klover.ai/nike-ai-strategy-analysis-of-dominance-in-athletic-apparel-footwear/. Secondary source; Accessed August 10, 2026.

Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Quarterly Journal of Economics 132, 665–712.

Krupa, T.J., Mullaney, M.S., 2026. The Use of Artificial Intelligence in Decision-Making: Evidence from the Effectiveness of Corporate Tax Strategies. Review of Accounting Studies 31, 704–744.

Lee, A., 2026. Expedia's AI Timeline Runs Through Its Talent Strategy. Skift. https://skift.com/2026/07/20/expedia-cto-ramana-thumu-ai-talent-strategy/.

Lev, B., Radhakrishnan, S., 2005. The Valuation of Organization Capital, in: Corrado, C., Haltiwanger, J., Sichel, D. (Eds.), Measuring Capital in the New Economy. University of Chicago Press, Chicago, IL, pp. 73–110.

Li, K., Mai, F., Shen, R., Yang, C., Zhang, T., 2026. Dissecting Corporate Culture Using Generative

AI. Review of Financial Studies 39, 253–296.

Lightcast, 2026a. Job Posting Analytics (JPA) Methodology. https://docs.lightcast.io/data/docs/job-posting-analytics-jpa-methodology. Accessed August 14, 2026.

Lightcast, 2026b. Lightcast Data: Basic Overview. https://docs.lightcast.io/data/docs/lightcast-data-basic-overview. Accessed August 14, 2026.

Liu, E., Ma, S., Veldkamp, L., 2025. Data Sales and Data Dilution. Journal of Financial Economics 169, 104053.

Lustig, H., Syverson, C., Van Nieuwerburgh, S., 2011. Technological Change and the Growing Inequality in Managerial Compensation. Journal of Financial Economics 99, 601–627.

March, J.G., 1991. Exploration and Exploitation in Organizational Learning. Organization Science 2, 71–87.

McElheran, K., Li, J.F., Brynjolfsson, E., Kroff, Z., Dinlersoz, E., Foster, L., Zolas, N., 2024. AI Adoption in America: Who, What, and Where. Journal of Economics & Management Strategy 33 (2), 375–415.

McElheran, K., Yang, M.J., Brynjolfsson, E., Kroff, Z., 2025. The Rise of Industrial AI in America: Microfoundations of the Productivity J-Curve(s). Working Paper.

McLemore, P., Mihov, A., 2026. AI and Operational Losses: Evidence from U.S. Bank Holding Companies. The Review of Corporate Finance Studies, cfag003.

Mihet, R., Philippon, T., 2019. The Economics of Big Data and Artificial Intelligence, in: Choi, J.J., Ozkan, B. (Eds.), Disruptive Innovation in Business and Finance in the Digital World. Emerald Publishing Limited. https://www.emerald.com/book/chapter-pdf/8928654/s1569-376720190000020006.pdf.

Mitsubishi UFJ Financial Group, 2024. Who We Are. https://www.tradepayables.com/mufg/who-we-are.html. Accessed June 10, 2026.

Morgan Stanley Wealth Management, 2024. Morgan Stanley Wealth Management Announces Latest Game-Changing Addition to Suite of GenAI Tools. https://www.morganstanley.com/press-releases/ai-at-morgan-stanley-debrief-launch.

New York Times, 2019. Google's Shadow Work Force: Temps Who Outnumber Full-Time Employees. https://www.nytimes.com/2019/05/28/technology/google-temp-workers.html.

NIKE, Inc., 2024. Creating the Unreal: How Nike Made Its Wildest Air Footwear Yet. Nike Newsroom. https://about.nike.com/en/stories/nike-design-athlete-imagined-revolution. Accessed August 10, 2026.

Nonaka, I., 1994. A Dynamic Theory of Organizational Knowledge Creation. Organization Science 5, 14–37.

Noy, S., Zhang, W., 2023. Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. Science 381, 187–192.

OECD, Eurostat, 2018. Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation. OECD Publishing.

Olavsrud, T., 2026. The Gen AI Helping Aetna Review Millions of Medical Records. CIO, July 31. https://www.cio.com/article/4202752/how-gen-ai-helps-aetna-review-millions-of-medical-records.html.

Olley, G.S., Pakes, A., 1996. The Dynamics of Productivity in the Telecommunications Equipment

Industry. Econometrica 64, 1263–1297.

Peters, R., Taylor, L.A., 2017. Intangible Capital and the Investment-q Relation. Journal of Financial Economics 123, 251–272.

Polanyi, M., 1966. The Tacit Dimension. Routledge and Kegan Paul, London.

Prescott, E.C., Visscher, M., 1980. Organization Capital. Journal of Political Economy 88, 446–461.

Raina, R., Madhavan, A., Ng, A.Y., 2009. Large-Scale Deep Unsupervised Learning Using Graphics Processors, in: Proceedings of the 26th Annual International Conference on Machine Learning (ICML), ACM, Montreal, Quebec, Canada. pp. 873–880.

Revelio Labs, 2026. Revelio Public Labor Statistics. https://www.reveliolabs.com/public-labor-statistics. Accessed August 14, 2026.

Rivkin, J.W., 2001. Reproducing Knowledge: Replication Without Imitation at Moderate Complexity. Organization Science 12, 274–293.

Roche Holding AG, 2024. Annual Report 2024. https://assets.roche.com/f/174029/x/74e4b5f7dc/roche-annual-report-2024.pdf. Accessed June 13, 2026.

Rock, D., 2019. Engineering Value: The Returns to Technological Talent and Investments in Artificial Intelligence. Working Paper.

Schubert, G., 2025. Organizational Technology Ladders: Remote Work and Generative AI Adoption. Working Paper.

Sheng, J., Sun, Z., Yang, B., Zhang, A.L., 2026. Generative AI and Asset Management. Review of Financial Studies, hhag050.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S.S., Wei, J., Chung, H.W., Scales, N., Tanwani, A., Cole-Lewis, H., Pfohl, S., Payne, P., Seneviratne, M., Gamble, P., Kelly, C., Babiker, A., Schärli, N., Chowdhery, A., Mansfield, P., Demner-Fushman, D., Agüera y Arcas, B., Webster, D., Corrado, G.S., Matias, Y., Chou, K., Gottweis, J., Tomasev, N., Liu, Y., Rajkomar, A., Barral, J., Semturs, C., Karthikesalingam, A., Natarajan, V., 2023. Large Language Models Encode Clinical Knowledge. Nature 620, 172–180.

Sohel, 2026. How Nike Is Using AI: Strategy, Tools, Real Examples. Dilogs.ai Blog. https://www.dilogs.ai/blog/how-nike-is-using-ai/.

Son, H., 2026. Morgan Stanley Will Soon Open Its Trillion-Dollar Wealth Management Funnel to AI Agents. CNBC, June 3. https://www.cnbc.com/2026/06/03/ai-agents-morgan-stanley-wealth-management-funnel.html.

Tambe, P., Hitt, L., Rock, D., Brynjolfsson, E., 2020. Digital Capital and Superstar Firms. NBER Working Paper 28285.

Team DigitalDefynd, 2026. 10 Ways Nike Is Using AI [Case Study]. DigitalDefynd. https://digitaldefynd.com/IQ/ways-nike-use-ai/.

The Home Depot, 2025a. The Home Depot Introduces Magic Apron, a Suite of Advanced AI Tools to Help Customers with Their Home Improvement Needs. Press release, March 6. https://ir.homedepot.com/news-releases/2025/03-06-2025-130241718.

The Home Depot, 2025b. Unveiling Magic Apron: The Home Depot's Smartest Tool Yet. Corporate newsroom, March 6. https://corporate.homedepot.com/news/company/unveiling-magic-apron-home-depots-smartest-tool-yet.

The Home Depot, 2026a. AI at the Home Depot: A Look at the Future of Home Improvement.

Corporate newsroom, March 30. https://corporate.homedepot.com/news/company/ai-home-depot-look-future-home-improvement.

The Home Depot, 2026b. The Home Depot and Google Cloud Launch Agentic AI Tools to Help Customers and Associates Bring Projects from 'How-to' to 'Done'. Press release, January 11. https://corporate.homedepot.com/news/partnerships/home-depot-and-google-cloud-launch-agentic-ai-tools-help-customers-and-associates.

Trammell, P., Korinek, A., 2026. Economic Growth Under Transformative AI. Annual Review of Economics 18, 589–611.

Turing, A.M., 1950. Computing Machinery and Intelligence. Mind 59, 433–460.

Veldkamp, L., Chung, C., 2024. Data and the Aggregate Economy. Journal of Economic Literature 62, 458–484.

Webb, M., 2020. The Impact of Artificial Intelligence on the Labor Market. Working Paper.

Woeppel, M., Yavuz, M.D., 2026. Measuring Firms' Capability to Absorb External Knowledge. Journal of Financial Economics 183, 104322.

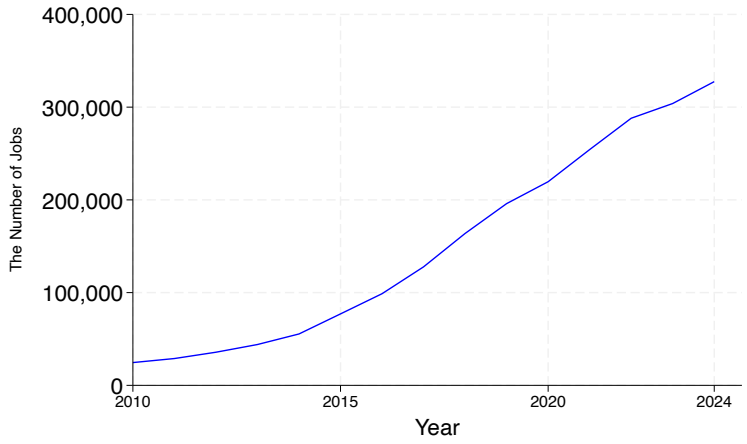
World Economic Forum, 2021. Factories of the Future Find Growth and Sustainability Through Digitalization. Press release, March 17. https://www.weforum.org/press/2021/03/factories-of-the-future-find-growth-and-sustainability-through-digitalization/.

World Economic Forum, 2023. Against Economic Headwinds, 18 Manufacturing Lighthouses Show How to Boost Productivity and Sustainability by Scaling Advanced Technologies Across Networks. Press release, January 13. https://www.weforum.org/press/2023/01/against-economic-headwinds-18-manufacturing-lighthouses-show-how-to-boost-productivity-and-sustainability-by-scaling-advanced-technologies-across-networks/.

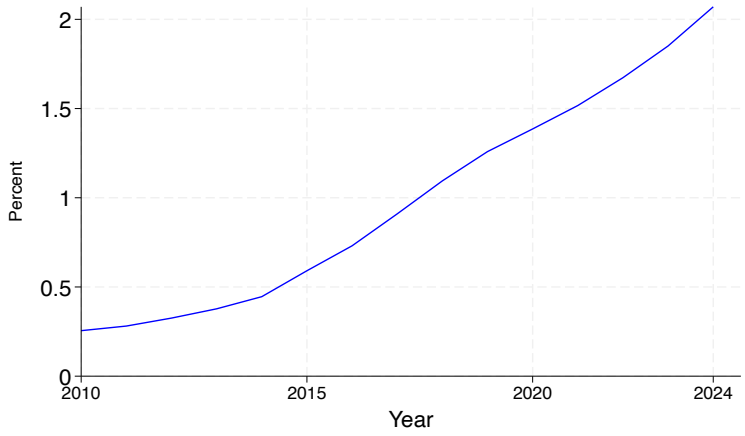
Yotzov, I., Barrero, J.M., Bloom, N., Bunn, P., Davis, S.J., Foster, K.M., Jalca, A., Meyer, B.H., Mizen, P., Navarrete, M.A., Smietanka, P., Thwaites, G., Wang, B.Z., 2026. Firm Data on AI. NBER Working Paper 34836.

Zander, U., Kogut, B., 1995. Knowledge and the Speed of the Transfer and Imitation of Organizational Capabilities: An Empirical Test. Organization Science 6, 76–92.

Figure 1. Time Series of AI Jobs



(a) Number of AI Jobs

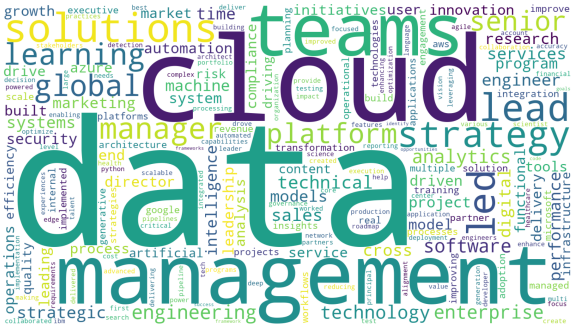


(b) Share of AI Jobs

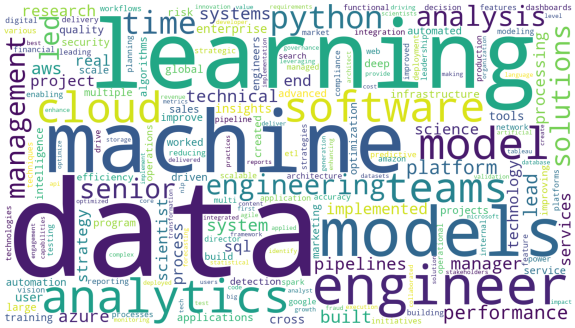
This figure plots annual AI-related jobs in the U.S. public-firm sample from 2010 to 2024. Panel (a) reports the total number of AI-related U.S. jobs. Panel (b) reports the aggregate AI job share, defined as AI-related U.S. jobs divided by total U.S. jobs in the same year, expressed as a percentage.

Figure 2. Word Clouds for AI Technology Subcategories

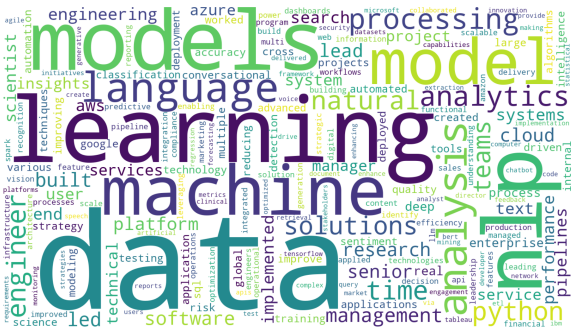
(a) General AI



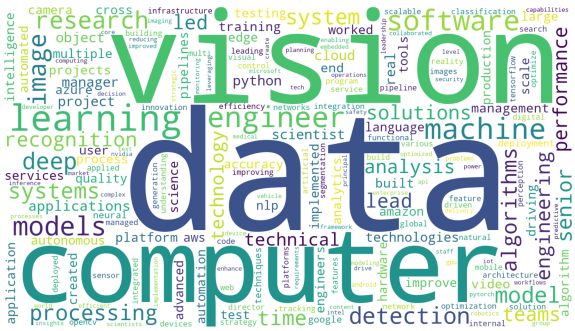
(b) ML



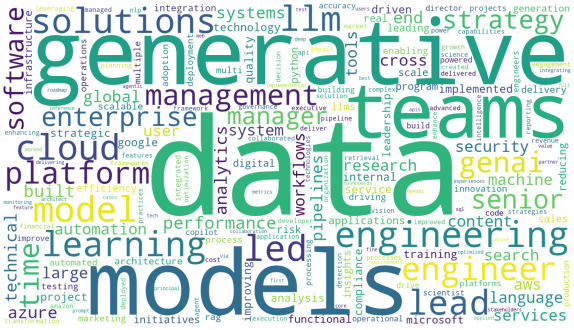
(c) NLP



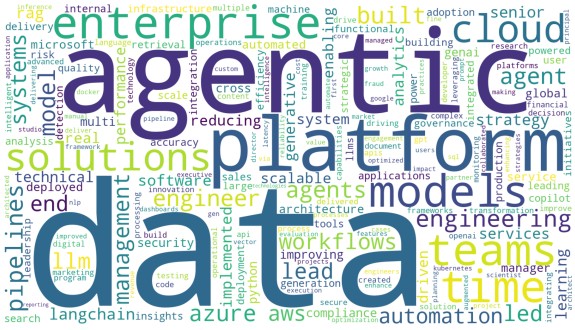
(d) CV



(e) Generative AI

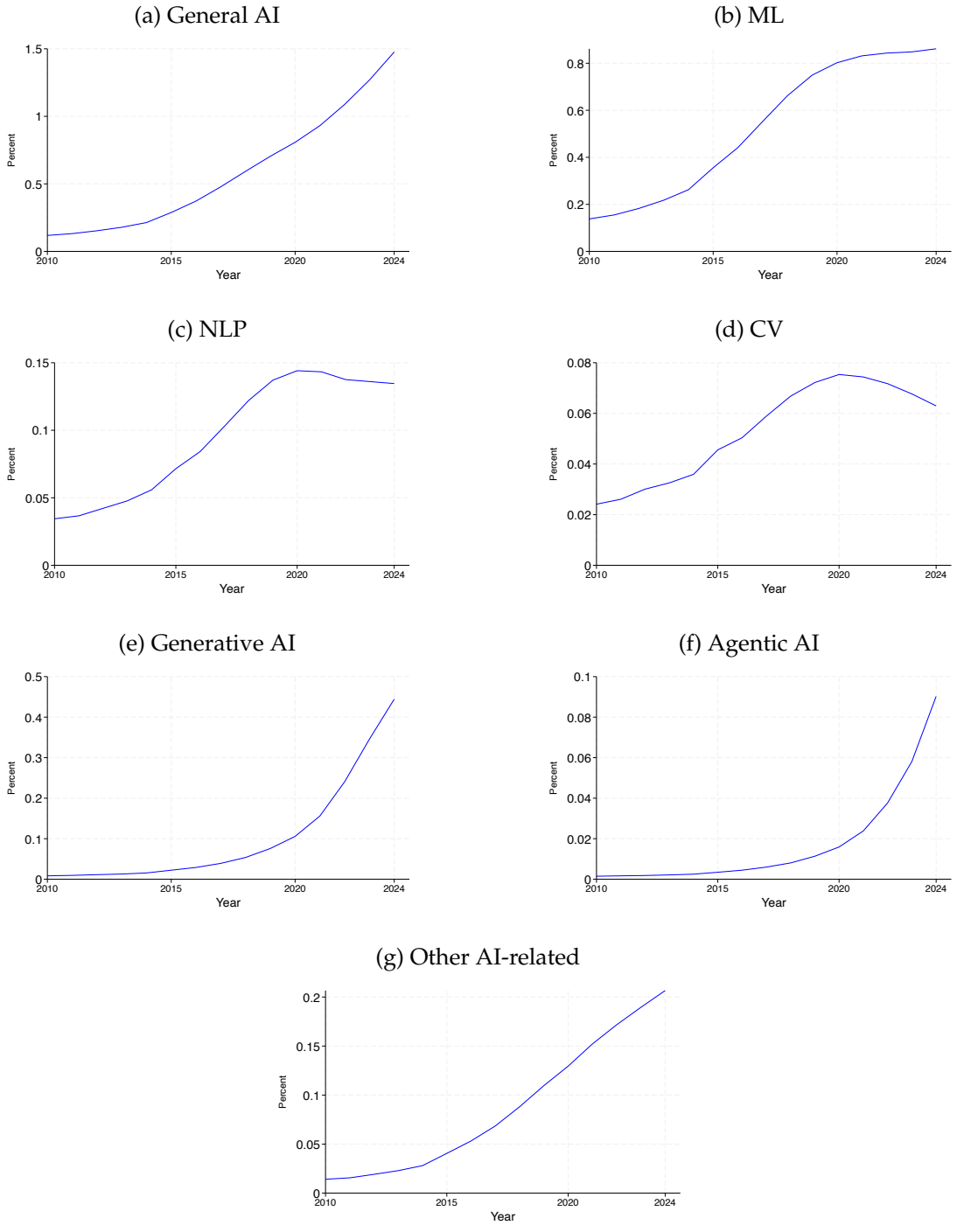


(f) Agentic AI



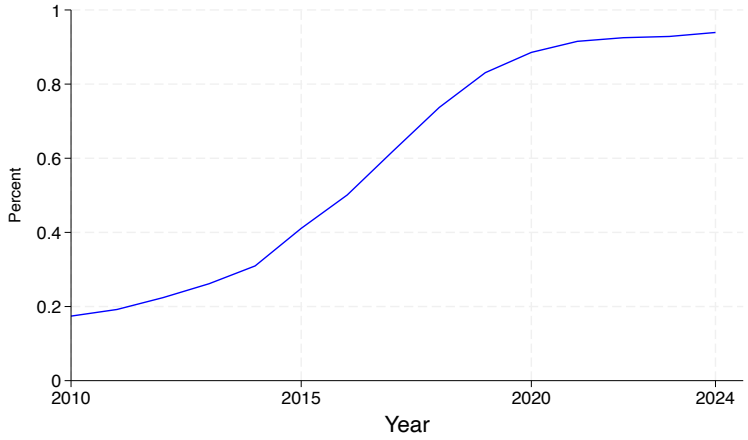
This figure plots word clouds for six AI technology subcategories: general AI, machine learning (ML), natural language processing (NLP), computer vision (CV), generative AI, and agentic AI. The word clouds are constructed from job titles and job descriptions of AI-related U.S. jobs in the U.S. public-firm sample from 2010 to 2024. For each technology subcategory, words are extracted from positions classified as related to that subcategory based on whether their job descriptions or titles contain keywords specific to that AI technology subcategory. The terms "AI" and "artificial intelligence" are excluded from the word clouds, as they appear in the vast majority of jobs in every subcategory and would otherwise dominate all panels.

Figure 3. Share of AI Jobs by AI Technology Subcategory

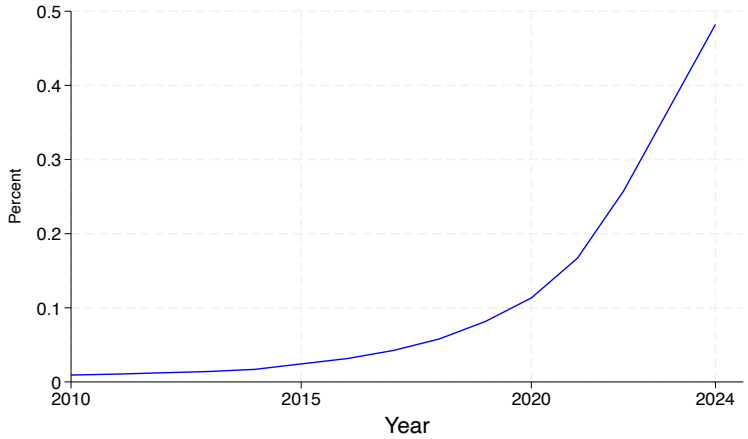


This figure plots annual AI-related U.S. job shares by AI technology subcategory in the U.S. public-firm sample from 2010 to 2024: general AI, machine learning (ML), natural language processing (NLP), computer vision (CV), generative AI, agentic AI, and other AI. The "Other AI-related" category includes AI-related jobs that are not assigned to the six named technology subcategories. Each panel reports the aggregate job share for each subcategory, defined as AI-related U.S. jobs in that subcategory divided by total U.S. jobs in the same year, expressed as a percentage. A job is classified into an AI subcategory if its job description or title contains keywords specific to that AI technology subcategory. Categories are not mutually exclusive, so a job can appear in multiple panels if it matches keywords from multiple AI technology subcategories.

Figure 4. Share of AI Jobs: Wave I vs. Wave II AI Technologies



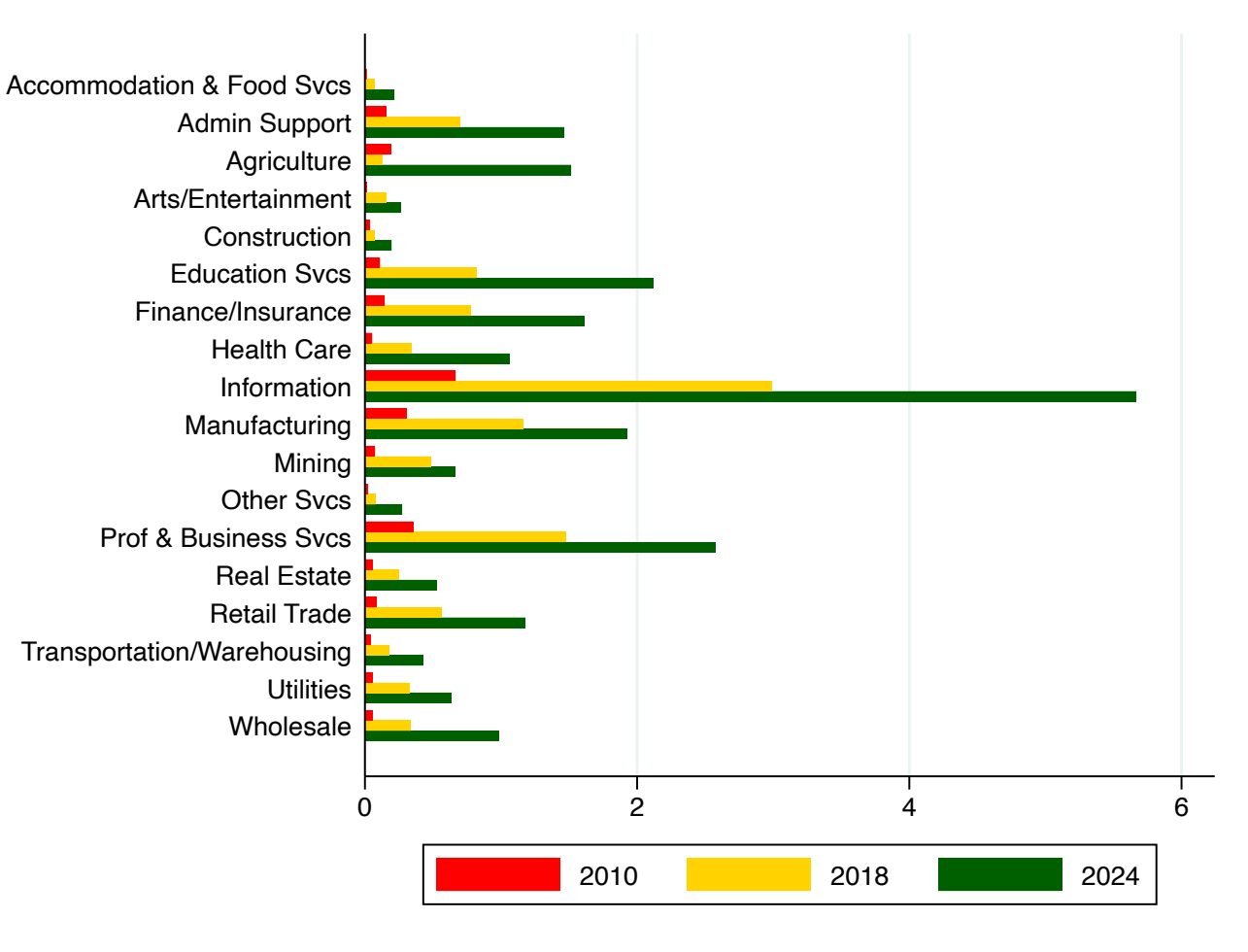
(a) Wave I: ML, NLP, and CV



(b) Wave II: Gen AI and Agentic AI

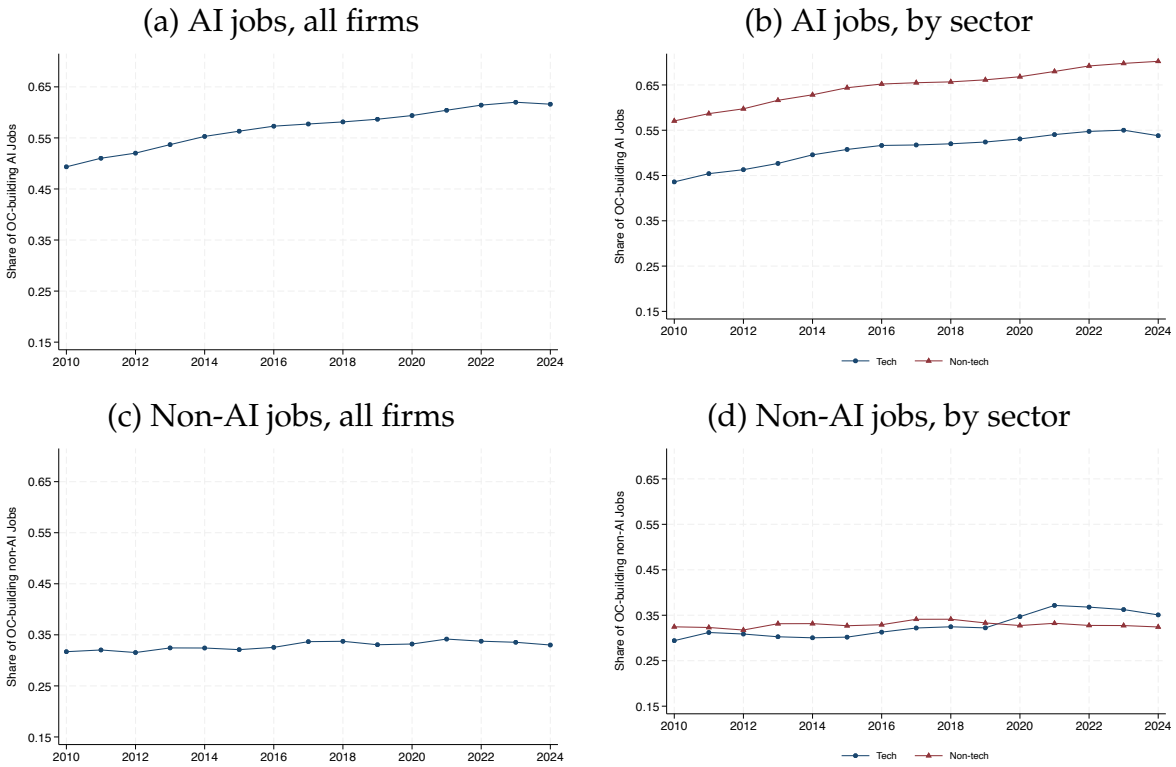
This figure plots the share of AI-related jobs in U.S. public firms from 2010 to 2024, separately for Wave I and Wave II AI technologies. Panel (a) reports the share of "Wave I" AI jobs, defined as jobs with job titles or descriptions containing at least one of the keywords from machine learning (ML), natural language processing (NLP), or computer vision (CV) technology subcategories. Panel (b) reports the share of "Wave II" AI jobs, defined as jobs with job titles or descriptions containing at least one of the keywords from generative AI (Gen AI) or agentic AI technology subcategories. Each share equals U.S. AI jobs in the indicated wave divided by all U.S. jobs in the same year, expressed as a percentage. Wave I and Wave II are not mutually exclusive.

Figure 5. Share of AI Jobs by Industry Sector



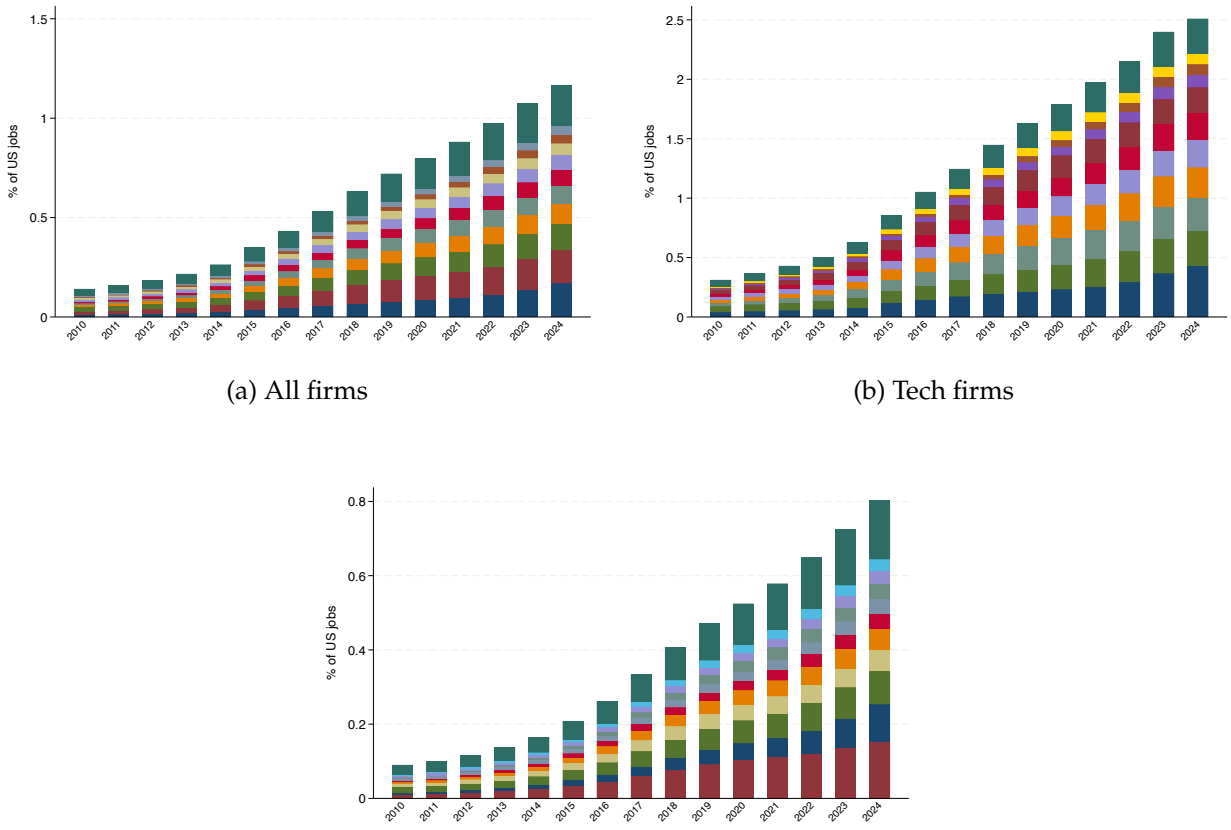
This figure presents the share of AI-related jobs at the industry level, based on the sample of U.S. public firms. For each sector (based on 2-digit NAICS industry codes) and year (2010, 2018, and 2024), the figure reports the ratio of the total number of AI-related U.S. jobs in that sector divided by the corresponding total number of U.S. jobs in that sector, expressed as a percentage.

Figure 6. Share of Organization-Capital-Building Jobs Among AI and Non-AI Jobs



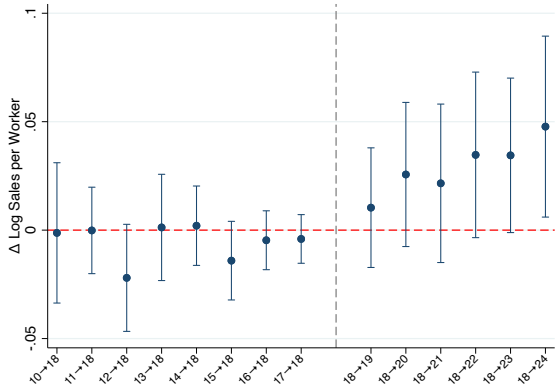
This figure plots the shares of organization-capital-building jobs among AI-related and non-AI-related jobs from 2010 to 2024 in the U.S. public-firm sample. Panel (a) plots the share of AI jobs that build organization capital, defined as the ratio of the number of AI-related jobs that build organization capital divided by the total number of AI-related jobs. Panel (b) plots the share of AI jobs that build organization capital separately for tech firms (2-digit NAICS 51, 54, and semiconductor manufacturing) and non-tech firms. Panel (c) plots the share of non-AI jobs that are classified as building organization capital, defined as the ratio of the number of non-AI-related jobs that build organization capital divided by the total number of non-AI-related jobs. Panel (d) plots the share of non-AI jobs that are organization-capital-building separately for tech and non-tech firms.

Figure 7. Organization Capital Categories for AI Jobs

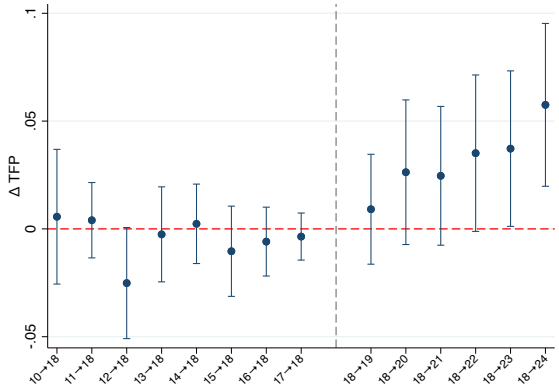


This figure plots the shares of AI-related jobs by organization-capital-building category from 2010 to 2024 in the U.S. public-firm sample. Each series reports the ratio of the number of organization-capital-building AI jobs assigned to each category divided by the total number of U.S. jobs, expressed in percentages. Panel (a) plots the shares for all firms. Panel (b) plots the shares for tech firms (2-digit NAICS 51, 54, and semiconductor manufacturing). Panel (c) plots the shares for non-tech firms. In each panel, we show the largest ten categories for that sample and group all remaining categories into "Other." Colors denote the same category in every panel.

Figure 8. AI Investments and Productivity Growth Over Time



(a) Log Sales per Worker



(b) Revenue TFP

This figure plots the association between firms' AI investment over 2018–2024 and productivity growth over different windows for U.S. public firms (in non-tech sectors). The independent variable is the standardized change in a firm's share of AI jobs from 2018 to 2024, ΔShare AI Jobs$_i^{18\to24}$, which is identical to the independent variable used in our main long-differences regressions for the 2018–2024 sample. Each point plots the coefficient $\hat{\beta}$ from a separate cross-sectional regression of growth in productivity, $\Delta Y_i^{T\to T'}$, on ΔShare AI Jobs$_i^{18\to24}$, controlling for 2-digit NAICS sector fixed effects and weighted by the number of jobs in Revelio in 2018. Panel (a) reports results for log sales per worker, and Panel (b) is for revenue total factor productivity (TFP). To the left of the dashed line, the dependent variable is $\Delta Y_i^{T\to2018}$ for $T = 2010, \ldots, 2017$. For example, the very first dot measures how changes in firm AI investments from 2018 to 2024 relate to ex-ante change in productivity from 2010 to 2018. The next dot shortens the period over which ex-ante productivity is measured by one year (from 2011 to 2018), with the last dot on the left using changes in productivity from 2017 to 2018. To the right of the dashed line, the dependent variable is $\Delta Y_i^{2018\to T}$ for $T = 2019, \ldots, 2024$: the very first dot after the dashed line measures how changes in firm AI investments from 2018 to 2024 relate to the change in productivity from 2018 to 2019, with the last dot using changes in productivity from 2018 to 2024. The vertical bars indicate 95% confidence intervals. Standard errors are clustered at the 5-digit NAICS industry level.

Table 1. Organization Capital Categories: Distribution of AI vs. Non-AI Jobs

Panel A: Builds organization capital

Reason category	AI jobs %	Non-AI jobs %
Reusable data platforms & pipelines	9.1	2.2
Reusable AI/ML platforms & MLOps	4.8	0.1
Firm-specific predictive & decision models	2.3	0.2
Proprietary-knowledge GenAI/NLP systems	6.9	0.4
Reusable software platforms & workflows	2.2	1.8
Cross-functional operating routines	6.3	4.8
Reusable cloud, infrastructure & reliability operations	3.4	2.0
Security, compliance & risk systems	1.9	1.4
Marketing & revenue-operations systems	4.1	1.9
Customer-service & support systems	2.0	0.8
Product & design systems	4.7	2.0
Manufacturing & process systems	1.6	3.5
Supply-chain & logistics systems	1.5	2.1
R&D & scientific platforms	1.4	0.5
Financial & risk decision systems	3.0	1.3
Hiring, training & talent development	1.4	5.8
Alliance & partnership systems	0.5	0.2
Knowledge organization & ontologies	0.4	0.0
Subtotal, builds organization capital	57.5	31.0

Panel B: Does not build organization capital

Reason category	AI jobs %	Non-AI jobs %
Bespoke client consulting/implementation	7.8	8.8
Portable analytics & one-off models	12.6	4.1
Broadly applicable research & algorithms	8.9	1.3
Customer-facing products & IP	3.8	3.9
Standardized/portable software work	3.3	2.4
Transactional sales, marketing & account management	2.9	7.6
Individual recruiting & hiring transactions	0.9	0.7
Standardized routine operations, support, maintenance	0.7	37.9
Project/program management (episodic)	0.3	0.5
Training & general education	0.5	0.7
Portable legal & contracting	0.8	1.1
Subtotal, does not build organization capital	42.5	69.0

This table reports the distribution of jobs by their organization capital categories for AI-related and non-AI-related U.S. jobs in the U.S. public-firm sample from 2010 to 2024. A job is classified as organization-capital-building and assigned to a standardized organization-capital category based on its job title and description. Panel A reports the percentages of AI and non-AI jobs in categories classified as building organization capital. Panel B reports the percentages of AI and non-AI jobs in categories classified as not building organization capital.

Table 2. AI Investments and Productivity Growth

Panel A: 2010–2018

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	-0.004	0.002	0.014	0.012
	(0.014)	(0.015)	(0.012)	(0.012)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.089	0.171	0.047	0.150
Observations	1,222	1,222	1,091	1,091

Panel B: 2018–2024

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.048**	0.062**	0.061***	0.075***
	(0.021)	(0.026)	(0.020)	(0.024)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.178	0.226	0.071	0.153
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments among U.S. public firms (in non-tech sectors). All changes in productivity and the share of AI jobs are measured from 2010 to 2018 in Panel A and from 2018 to 2024 in Panel B. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue total factor productivity (TFP) (columns 3 and 4). The independent variable is growth in the share of AI-related jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of each period (2010 in Panel A and 2018 in Panel B): firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the sample period (2010 in Panel A and 2018 in Panel B). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3. Investments in Different Waves of AI Technologies and Productivity Growth

	Δ Log Sales per Worker						Δ TFP					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δ Share Wave I Only AI Jobs	0.057**	0.065**					0.062**	0.069**				
	(0.024)	(0.028)					(0.024)	(0.027)				
Δ Share Wave II Only AI Jobs			0.034*	0.040**					0.058***	0.067***		
			(0.017)	(0.018)					(0.014)	(0.018)		
Δ Share Wave I & II AI Jobs					0.040**	0.056**					0.052***	0.070***
					(0.018)	(0.024)					(0.018)	(0.024)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Adj R-Squared	0.184	0.230	0.161	0.203	0.181	0.230	0.068	0.145	0.051	0.127	0.078	0.165
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,173	1,173	1,173	1,173	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on contemporaneous firm-level changes in share of AI jobs related to different waves of AI technologies. The sample is U.S. public firms in non-tech sectors. All changes in productivity and the share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1–6) and changes in revenue total factor productivity (TFP) (columns 7–12). In columns 1, 2, 7, and 8, the independent variable is the change in the AI share of jobs with only Wave I AI (machine learning, natural language processing, or computer vision) skills. In columns 3, 4, 9, and 10, the independent variable is the change in the AI share of jobs with only Wave II AI (generative or agentic AI) skills. In columns 5, 6, 11, and 12, the independent variable is the change in the AI share of jobs with both waves of AI skills. The AI shares are the ratio of the AI job count in each AI job category (only Wave I, only Wave II, or both) normalized by firms' total U.S.-based jobs. In all columns, the independent variables are standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 4, 6, 8, 10, and 12 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the sample period in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4. AI Investments and Productivity Growth: IV Estimates

	Δ Log Sales per Worker			Δ TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	0.160	0.158**	0.144***	0.135*	0.110**	0.114***
	(0.098)	(0.062)	(0.048)	(0.076)	(0.048)	(0.041)
Industry FE	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
State FE	N	Y	Y	N	Y	Y
Control Pre-trend	N	N	Y	N	N	Y
F Statistic	10.1	17.9	26.2	10.5	18.8	26.0
Observations	633	627	531	576	570	528

This table shows 2SLS estimates of the effect of changes in AI investments on changes in firm productivity for U.S. public firms (in non-tech sectors). All changes in productivity and the share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1–3) and changes in revenue total factor productivity (TFP) (columns 4–6). The (instrumented) independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. The instrument measures ex-ante firm-level exposure to the supply of AI-trained university graduates from universities with historically strong AI research: it is the share of STEM workers in the firm between 2005 and 2010 who graduated from ex-ante AI-strong research universities. The first-stage F-statistics of the instrument are reported for all specifications. All specifications control for industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 20 universities. All columns also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 2, 3, 5, and 6 add state fixed effects, and columns 3 and 6 additionally control for firm-level changes in log sales, log employment, and TFP from 2010 to 2018. Regressions are weighted by the number of jobs in Revelio at the start of the sample period in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5. AI Investments and Organization Capital Growth

Panel A: 2010–2018

	Δ Log(Organization Capital/PP&E)		Δ Log(Organization Capital/Total Assets)	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	-0.053	-0.042	0.034	-0.002
	(0.037)	(0.046)	(0.022)	(0.034)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.104	0.209	0.171	0.262
Observations	851	851	1,028	1,028

Panel B: 2018–2024

	Δ Log(Organization Capital/PP&E)		Δ Log(Organization Capital/Total Assets)	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.104***	0.089***	0.119***	0.110***
	(0.028)	(0.027)	(0.043)	(0.036)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.095	0.124	0.092	0.126
Observations	886	886	1,105	1,105

This table reports the coefficients from long-differences regressions of changes in organization capital (based on capitalized SG&A) on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in organization capital and the share of AI jobs are measured from 2010 to 2018 in Panel A and from 2018 to 2024 in Panel B. The dependent variables are changes in the natural log of organization capital normalized by PP&E in columns 1 and 2 and changes in the natural log of organization capital normalized by total assets in columns 3 and 4. The main independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of each period (2010 in Panel A and 2018 in Panel B): firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the sample period (2010 in Panel A and 2018 in Panel B). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 6. Organization-Capital-Building AI Jobs and Productivity Growth

	Δ Log Sales per Worker			Δ TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.069***	0.085***	0.084***	0.081***	0.100***	0.094***
	(0.025)	(0.028)	(0.026)	(0.025)	(0.027)	(0.025)
Δ Share of Non-OC-building AI Jobs			0.002			0.014
			(0.021)			(0.022)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.188	0.238	0.237	0.078	0.167	0.167
Observations	1,306	1,306	1,306	1,163	1,163	1,163

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI jobs that build (OC-building) and do not build (Non-OC-building) organization capital (OC) in U.S. public firms (in non-tech sectors). All changes in productivity and the share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1–3) and changes in revenue total factor productivity (TFP) (columns 4–6). The independent variables are growth in the share of OC-building AI jobs in all columns, and the share of non-OC-building AI jobs in columns 3 and 6, which are both normalized by firms' total U.S.-based jobs and standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 3, 5, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the sample period in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 7. Heterogeneous Relationship between AI Investments and Productivity Growth by Initial Organization Capital

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs*Below-Median OC	0.070*	0.092**	0.088**	0.105***
	(0.043)	(0.037)	(0.039)	(0.036)
Δ Share AI Jobs*Above-Median OC	0.034	0.050	0.027	0.052
	(0.038)	(0.044)	(0.033)	(0.039)
NAICS2*Size tercile*Above-Median OC FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.200	0.259	0.184	0.245
Observations	875	875	836	836

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), separately for firms with above-median and below-median initial levels of organization capital (OC), based on capitalized SG&A. All changes in productivity and the share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue total factor productivity (TFP) (columns 3 and 4). Firms within each 2-digit NAICS industry sector are split at the median level of the ratio of organization capital (based on capitalized SG&A) normalized by PP&E in 2018 into two groups, "Above-Median OC" and "Below-Median OC." The main independent variable is growth in the share of AI jobs, standardized to mean zero and standard deviation of one, interacted with an indicator for above- versus below-median initial organization capital. All specifications control for fixed effects for each 2-digit NAICS industry × firm size tercile × above-median organization capital cell. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the sample period in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix

A Data and Measurement of AI Investment and Productivity 59
A.1 Linking Revelio Labs Firms to Compustat 59
A.2 Construction of Firm-Level AI Measures 66
A.3 Construction of AI Job Categories 81
A.4 Measures of Productivity 86

B Organization Capital: Definition, Measurement, and Case Studies 133
B.1 Definition and Neighboring Concepts 133
B.2 Measuring OC Using Job Descriptions 137
B.3 Case Studies of AI Building Organization Capital 145

C Robustness and Additional Results 187

D Instrument Construction and Validation 236

E A Model of AI and Organization Capital 248
E.1 Model Setup 249
E.2 AI as a Codification Technology 254
E.3 Event-Study Dynamics 257
E.4 Endogenous Adoption over the Life Cycle 260
E.5 Taking the Model to Data 262
E.6 Extension: OC as a Substitute for Measured Inputs 263

A. Data and Measurement of AI Investment and Productivity

A.1. Linking Revelio Labs Firms to Compustat

This section describes how we link each firm in Revelio Labs to Compustat over time. The resulting mapping is many-to-one in each year, since multiple Revelio firm identifiers (`rcid`) can map to the same Compustat firm (`gvkey`).[41] We construct our own mapping rather than use Revelio's direct linkage so that it can vary over time and accommodate mergers, acquisitions, and changes in corporate structure that alter firm identifiers.[42]

The procedure has five stages: (i) four direct matching methods that exploit the firm identifiers Revelio reports for each `rcid`; (ii) extension of these matches from parent firms to subsidiaries using Revelio's dynamic parent-subsidiary panel; (iii) consolidation of all candidate matches; (iv) a three-step false-positive screen that combines algorithmic similarity scoring, a large-language-model audit, and human review; and (v) production of the final panel.

A.1.1. Data inputs

(i) **Compustat firm-year master:** For each Compustat firm-year, we extract the firm identifier (`gvkey`), name, headquarters address (ZIP, city, state), NAICS code, IPO date, and other identifiers (`cusip`, `cik`, and `ein`). We drop firm-year observations with missing or zero employment.

(ii) **Revelio firm mapping file:** For each firm in Revelio Labs, we observe Revelio's firm identifier (`rcid`), firm name, ultimate-parent firm identifier (`ultimate_parent_rcid`) and name, headquarters location, founding year, NAICS code, and firm identifiers linked by Revelio, including `gvkey`, `cusip`, `cik`, and `ein`.

(iii) **Revelio employment:** Annual count of U.S. and global employed workers associated with each firm in Revelio Labs.

(iv) **Subsidiary structure data:** Revelio's panel of parent-subsidiary relationships with effective start and end dates, allowing us to identify which `rcid` is owned by which ultimate parent in any given year.

(v) **Historical identifier crosswalks from WRDS:** We use three WRDS link tables, each indexed by effective start and end dates, to build the historical mappings we need.

- `ccmxpf_lnkhist` – the CRSP-Compustat Merged link history, which gives the historical mapping between `gvkey` and CRSP `permno` with link start (`linkdt`) and end (`linkenddt`) dates.
- `msenames` – the CRSP stock event names file, which gives the historical mapping between CRSP `permno` and `cusip` with effective name start (`namedt`) and end (`nameendt`) dates.

[41] Multiple Revelio firms sharing one `gvkey` are subsidiaries of a corporate parent.

[42] Examples include spin-offs, divestitures, reorganization, and reassignments of subsidiaries to a different parent.

- `wciklink_gvkey` – the WRDS SEC `cik` link table, which gives the historical mapping between `cik` and `gvkey` with link start (`link_start_date`) and end (`link_end_date`) dates.

A.1.2. Four direct matching methods

For each `rcid`, Revelio reports other commonly used firm identifiers, such as `gvkey`, `cusip`, `cik`, and `ein`. These identifiers are also available for firms in Compustat. For each `rcid` in Revelio, we attempt four matches in parallel to firms in Compustat using these identifiers from both datasets, each producing a candidate `gvkey`:

(i) **Direct `gvkey` match:** Revelio assigns a `gvkey` to a subset of `rcids` based on its own internal matching. We take this assignment as the first candidate.

(ii) **Historical CUSIP:** Using the Compustat-CRSP link, we construct a year-indexed panel of all `gvkey-cusip` pairs that are linked in each year. In the year where the matched cusip for a gvkey switches, we use the new cusip for that year. For each Revelio firm-year, we match the `rcid`'s `cusip` to this panel. If the same `cusip` is linked to a `gvkey` in year `t`, that `gvkey` becomes a candidate.

(iii) **Historical CIK:** The same logic as above is applied to the `gvkey-cik` panel from `wciklink_gvkey`.

(iv) **EIN:** We match the `rcid`'s `ein` directly to the Compustat firm's `ein`. Unlike CUSIP and CIK, there is no widely available historical EIN-to-gvkey crosswalk, so we use Compustat's current EIN. This is the least reliable of the four methods for that reason.

Each method labels its successful matches with a tag (`match_on` $\in$ {`gvkey`, `cusip`, `cik`, `ein`}). A single `rcid` may be matched by more than one method, and to a `gvkey` different from another method's candidate in this step. We reconcile these conflicts in Section A.1.4.

A.1.3. Matching subsidiaries through their parent

Many Revelio firms are corporate subsidiaries that do not carry the parent's identifiers, so Section A.1.2 misses them. For example, the top-level Alphabet `rcid` carries Alphabet's `cusip` and `cik`, so Section A.1.2 matches it directly to Alphabet's `gvkey`. But some of Alphabet's subsidiary `rcids`, such as YouTube, DeepMind, and so on, carry no Compustat identifier of their own in Revelio's company mapping file. In addition, the company mapping file is static and does not provide information on the timing of the acquisitions. To pick up the subsidiaries dynamically, we use Revelio's parent-subsidiary panel together with Section A.1.2.

For every (`rcid`, year) cell, we look up the ultimate parent `rcid` active in that year. If the parent has been matched in Section A.1.2 and the ultimate parent's rcid is different from the firm's own rcid, we assign the parent's `gvkey` to the subsidiary for the same year and flag the match as one made through the parent (`subsidiary` $= 1$).

A.1.4. Consolidation

We append all direct matches from the above steps. When a single (rcid, year) cell has multiple direct matches that disagree (this is less than 1% of cases), we keep the most reliable one in the order gvkey > cusip > cik > ein.

We then match each rcid-year observation to the gvkey of the ultimate parent. We replace the gvkey that is matched directly with the gvkey of the ultimate parent after acquisition. We also replace the gvkey as missing if the rcid is matched to the ultimate parent's gvkey via direct matching before the acquisition.

The output of this stage is a panel of candidate matches at the (rcid, year) level: each row identifies a candidate gvkey, a match_on tag indicating which method produced the match, and a subsidiary indicator for whether the match flows through the parent company.

A.1.5. False-positive screen

Identifier-based matches can still be wrong. Two firms may share a historical cusip or cik after a corporate reorganization, or Revelio's ein for a subsidiary may belong to a holding company that is itself listed under a different name. To remove these false positives, we apply three screens in sequence.

Step 1: Similarity scoring. For each candidate (rcid, gvkey) match, we compute several similarity indicators between the Revelio firm and the Compustat firm:

- *Industry agreement* (NAICS match): equal to one if Revelio's 4-digit NAICS code and Compustat's 4-digit NAICS code match.
- *Location agreement* (state, city, ZIP): three separate binary indicators that equal one if the Revelio firm's headquarters state matches Compustat's, if the headquarters city matches, and if the headquarters ZIP code matches, respectively.
- *Name similarity*: a normalized string-similarity score between Revelio's company name and Compustat's company name.
- *Size ratio*: Revelio's annual employment count divided by Compustat's annual employment headcount.

A non-subsidiary candidate match is flagged as a suspected false positive if (i) the 4-digit NAICS codes disagree, the state, city, and ZIP code of the headquarters all disagree, and the name-similarity score is below 0.9; or (ii) Revelio's annual employment count for the rcid is more than twice Compustat's reported headcount for the gvkey; or (iii) the founding year in Revelio is after the IPO date in Compustat. We then review all potential false positives with over 50 employees in both Compustat and Revelio. For subsidiaries, we review all firms that have over 50 employees in both Compustat and Revelio and whose Revelio employment is over 5% of Compustat employment.

Step 2: Large language model audit. For each match flagged as suspect in Step 1, we ask the LLM (`Gemini-2.5 Pro`) whether the Revelio firm and the Compustat firm represent the same business entity in the relevant year.[43] The prompt supplies the model with both firms' names, headquarters, industries, employment levels, IPO dates, and a brief description, and asks for a yes-or-no judgment together with a short reason. We run this audit separately for non-subsidiary candidates where Revelio matches a public firm directly, and for subsidiary candidates where Revelio matches through an ultimate parent. The prompt is as follows:

```
You are an expert financial analyst. Your task is to determine whether two
company names refer to the SAME consolidated legal entity in a GIVEN year.

Your response MUST be a JSON object with two keys:

* "FP_indicator": A boolean value. Set to false if they are the same firm,
and true if they are different.

* "reason": A concise, one-sentence justification for your decision,
referencing the specific year if applicable.

Criteria for SAME firm (FP_indicator: false):

1. One firm acquired, merged with, or officially changed its name to the
other in or before the given year.

2. One firm is the legal parent company and the other is its legal subsidiary
in the given year.

Criteria for DIFFERENT firms (FP_indicator: true):

1. The firms are unrelated.

2. One name is merely a product or brand name, such as "Windows," and not a
distinct legal company.

3. The merger, acquisition, name change, or parent-subsidiary relationship
was established after the given year.
```

Step 3: Human review of negative LLM judgment. For every match that the language model labels "no", we conduct a manual review using public sources such as firm websites, SEC filings, and press coverage. This step guards against false negatives from the model. If we confirm the model's rejection, we drop the match. If we overrule it, we retain the match.

[43] We use Gemini-2.5 Pro because of its ability to draw on Google Search when assessing company relationships.

Final decision rule. After the three screening steps, we keep a candidate match if either (i) it has never been flagged as a suspect in Step 1, or (ii) it has been flagged but cleared by the LLM in Step 2 and has not subsequently been discarded by the human reviewer in Step 3. We drop the match if the human reviewer in Step 3 confirms it is a false-positive, and add back any match that the LLM rejected in Step 2 but whose rejection the human reviewer overrules in Step 3.

A.1.6. Final panel

The output of the procedure is an annual panel with one row per (`rcid`, `year`) cell, containing the matched Compustat `gvkey`, a `match_on` tag indicating which method produced the match (`gvkey`, `cusip`, `cik`, or `ein`), a `subsidiary` indicator equal to one if the match is made through the ultimate parent and zero if direct, and Revelio's annual employment count for the `rcid`.

A.1.7. Resume data employment coverage

Building on the identifier-level mapping panel above, we examine how firms' job counts in Revelio compare to the employment firms report in Compustat from 2010 through 2024. It is worth noting that the employment levels in the two datasets do not have to be equal: Compustat firms typically do not include temp workers, vendors, or contractors as part of their employee counts, while these workers might associate their employer with the firm that contracts with them directly or their direct employer. For example, Google is well known to have a two-tier worker system.[44] Representative surveys show that among workers responding that they are employees, 5 to 8% are actually independent contractors (Abraham et al., 2024, 2026a,b). Hence, for some firms Revelio job counts can, in principle, exceed those in Compustat.

Since Compustat only provides global firm-level employment (U.S.- and non-U.S.-based workers), in Revelio data we similarly aggregate employees located in all countries to the firm-year level. Figure A.1 sums employment each year across all matched firms: the blue line reports total jobs in Revelio (left y-axis) and the red line in Compustat (right y-axis). Two observations emerge. First, the two lines follow each other pretty closely over time. Second, global employment coverage improves over time: from 31% in 2010 to 53% in 2024.

The global employment coverage is lower than what Revelio reports for the U.S.-based workforce (about 2/3 of all U.S. workers are in Revelio data according to Revelio Labs): this is potentially due to lower usage of the LinkedIn website—the core source of online resumes—outside of the United States (e.g., few workers in China use LinkedIn). To investigate this hypothesis, we examine the employment coverage ratios by two firm characteristics: firms that are headquartered in the U.S. (those likely with larger U.S.-based employment) versus non-U.S. headquartered firms, and firms with the majority of workers being based in the U.S. in Revelio data (versus minority). Table A.1 shows that, on average, U.S.-based firms have about 20 percentage points higher coverage rates by both characteristics.[45]

[44] See New York Times (2019) and Guardian (2018).

[45] We manually researched a few companies with headquarters outside of the U.S. but with high shares of

Table A.2 complements Figure A.1 by comparing the firm-level distribution of Compustat employment and Revelio job counts in 2010, 2018, and 2024. Revelio job counts are lower than Compustat employment throughout the distribution, as expected given incomplete online-resume coverage, but as Figure A.1 shows the gap narrows over time to about 45% in 2018 and 53% in 2024.[46] For example, median Revelio jobs increase from 353 in 2010 to 644 in 2024, while median Compustat employment is roughly stable at about 1,165. The improvement is also visible in the upper tail, where the 95th percentile of Revelio jobs rises from 16,647 in 2010 to 34,207 in 2024. These firm-level patterns confirm that the aggregate improvement in employment coverage in Revelio in Figure A.1 is not driven only by a small number of very large firms.

We next examine the distribution of employment coverage by industry. Firms' industry codes are from Compustat. First, we look at the employment-weighted coverage: Figure A.2 reports statistics on the ratio of the sum of job counts across all firms in Revelio to the sum of employment across all firms in Compustat within each 2-digit NAICS industry-year cell.[47] The figure shows substantial industry variation in employment coverage, with relatively low coverage in sectors such as retail, wholesale, and accommodation & food services, where online-resume data under-represent lower-paid or less professionalized workers. Sectors with more high-skilled jobs, such as tech sectors, finance, or real estate, have the highest coverage ratios.

Second, we look at the equal-weighted employment coverage by industry: Figure A.3 reports, for each industry, the mean and median of firm-year coverage ratio pooled over 2010–2024. Specifically, we first compute the coverage ratio for each firm-year as total global Revelio jobs divided by Compustat employment, winsorize this ratio within year at the 1st and 99th percentiles, and then compute the mean and median across all firm-year observations within each industry. The equal-weighted coverage by industry is overall higher than the employment-weighted coverage across most industries. This is because employment is highly concentrated among large firms, while smaller firm-years tend to have relatively higher coverage ratios.[48]

We next examine the distribution of firm-level employment coverage ratios over time, defined as Revelio jobs divided by Compustat employment for each firm in each year. Figure A.4 shows the mean (in blue), the median (in red), and the interquartile range (gray shade) of the coverage ratio variable. Consistent with the improving aggregate coverage

employees in the U.S. in resume data. These cases confirm that these types of firms tend to have low coverage outside of the U.S. A few examples follow. "MITSUBISHI UFJ FINANCIAL GRP" (headquartered in Japan, but has more than 90% of employees in the U.S. according to Revelio data) reports in 2024 about 120,000 employees globally with the U.S. division having 13,000 employees, suggesting its share of U.S. employment is only 11% (Mitsubishi UFJ Financial Group, 2024). Similar discrepancies appear for Roche (headquartered in Switzerland): Revelio shows its U.S. share of employees is 97%, whereas public disclosures imply the U.S. employee share is no more than 26% (Roche Holding AG, 2024).

[46] We define the coverage in year t as total global Revelio jobs summed across all matched U.S. public firms in that year divided by total Compustat employment summed across the same firms.

[47] Bars report the mean and the median of the annual industry-level coverage ratio over the sample period from 2010 through 2024.

[48] The largest 10% of firm-years account for about 77% of total employment in our Compustat matched sample.

over time in Figure A.1, this figure shows that, on average, firm-level coverage goes up from 61% in 2010 to 94% in 2024. The figure also shows that there is a wide distribution of employment coverage across firms: at the end of the sample this ratio ranges from 33% at the 25th percentile to 135% at the 75th percentile. We dig into firms for which Revelio employment count exceeds Compustat's reported employee headcount. Oftentimes these are predominantly gig-economy firms (e.g., ride-share, food-delivery) or firms that tend to hire subcontractors (e.g., tech firms) for which Revelio captures independent contractors that are not typically included in Compustat employment.

Given our main empirical specification uses within-firm variation (we examine *changes* in firm outcomes), we next investigate whether firms with high or low employment coverage have issues tracking firm-level employment changes over time (i.e., firms for which changes in employment in Compustat and Revelio are not closely correlated). We partition the distribution of this employment coverage ratio into six intervals: below 0.05, 0.05–0.2, 0.2–1, 1–2, 2–3, and at least 3.[49] The two tail intervals identify observations for which the employment comparison is the least informative, since in these ranges Revelio coverage is either nearly zero or exceeds Compustat by more than a factor of three. The remaining intervals distinguish low but nontrivial coverage, partial coverage, approximate parity, and moderate over-coverage. We then compare Revelio and Compustat employment within each coverage interval. In each employment coverage interval, Appendix Table A.3 reports the corresponding correlations in levels and after absorbing firm fixed effects. For firm-years with coverage ratios between 0.2 and 3, the cross-sectional correlations are 0.90, 0.99, and 0.998 in the 0.2–1, 1–2, and 2–3 intervals, respectively, and the within-firm correlations are 0.87, 0.94, and 0.99. These patterns suggest that moderate under-coverage and moderate over-coverage mainly reflect level differences in measured employment rather than a failure of Revelio to track firm-level changes over time reported in Compustat.

In the lower tail where the coverage ratio is below 0.05, the within-firm correlation falls to 0.33, indicating that Revelio does not capture a significant share of the variation in changes in firm employment reported in Compustat. The 0.05–0.2 interval is borderline with a high cross-sectional correlation of 0.97 but a more modest within-firm correlation of 0.47. In the extreme upper tail (where coverage is at least 3) Revelio records substantially more workers than Compustat and the within-firm correlation is only 0.38. This analysis suggests that firms in the two extreme tails do not track changes in firm-level employment well in Compustat.

Further analysis shows that firms with low employee counts (either in Compustat or Revelio) also tend to have a low correlation in employment in the two datasets. Appendix Table A.4 shows that for firms with fewer than 50 jobs in Revelio, the cross-sectional correlation between employment in Compustat and job counts in Revelio is -1.6% and the within-firm correlation is 5.6%. For firms with fewer than 50 employees in Compustat, we similarly observe low cross-sectional correlation of 3.7%, while within-firm correlation rises to 50.5%.

Informed by these analyses, our baseline firm-level sample for our main regression

[49]The cutoff of 3 corresponds approximately to the 99th percentile of the coverage ratio distribution.

analysis retains firm-years with the employment coverage ratios between 0.05 and 3,[50] requires at least 50 workers in both Revelio and Compustat, and includes firms headquartered in the U.S. and having at least 50% of their workers in Revelio located in the U.S. For this sample of firms, we focus on the U.S.-based jobs to define AI-related and non-AI-related jobs. Table A.5 reports how the sample changes under each filter from 2010 to 2024. The matched Compustat–Revelio sample contains 4,783 firms in 2010 and 4,562 firms in 2024.[51] The employment coverage ratio filter leaves 4,099 firms in 2010 and 4,141 firms in 2024. Among the other one-at-a-time filters, the U.S.-headquarters, Revelio U.S. employment, U.S.-majority-worker, and Compustat employment filters leave 4,040, 3,309, 3,563, and 4,264 firms in 2010, respectively. In 2024, the corresponding counts are 3,487, 3,295, 2,997, and 3,991. Applying all these filters jointly leaves 2,482 firms in 2010 and 2,179 firms in 2024. It is worth noting that our descriptive evidence on AI-related jobs provided below focuses on a broader sample of U.S. public firms with the only restriction of using firm-years with the employment coverage ratios between 0.05 and 3. This allows us to examine statistics over a broader sample of jobs and re-examine some of the data issues we identified in this section.

A.2. Construction of Firm-Level AI Measures

In this section, we describe our methodology for identifying AI-related jobs in resume data: how we construct our AI-related keyword list and use it to measure the share of AI workers in each firm in our sample. Specifically, we proceed in several steps. We first build a list of AI-relevant skills from job postings in Lightcast data. We then expand that AI-relevant skill list using rule-based methods and an LLM. We next screen candidate AI keywords for false-positives and remove terms that tend to pick up AI-irrelevant jobs. Finally, we search the terms from the validated AI keyword list in resume data from Revelio to identify AI-relevant jobs.

We use this methodology because it has several key advantages. First, this approach for identifying AI-related jobs is 100% reproducible, which is critical in academic research. An alternative way is to ask an LLM to identify AI-related jobs, but then one needs to use judgment on how to define AI-related jobs (which is not trivial in quickly evolving AI technologies) and rely on nondeterministic LLM output (there is still variance in LLM classifications even if the LLM parameter to reduce variance—the temperature—is set to zero). Second, this approach is very inexpensive: one can run the relevant code on a laptop in an afternoon. In contrast, using a frontier LLM to classify billions of job records as AI-related or not would cost tens of thousands of dollars. Third, this approach is data-driven: it allows us to uncover AI-relevant job terms in a quickly evolving and dynamic AI field, and as we show, is able to identify jobs across different waves of AI, such as natural language processing and computer vision jobs early on and generative and agentic AI jobs most recently. Most importantly, we show in this extensive appendix that this

[50] We keep the 0.05–0.2 interval because it continues to align strongly with Compustat in the cross section, and the 1–3 range because moderate over-coverage reflects a systematic level difference rather than poor correlation of within-firm employment changes.

[51] The matched sample in Table C.2 contains fewer firms because the AI-share calculation requires firms to have at least one U.S. job.

methodology works well: it mostly picks up jobs involved in internal AI development with a false positive rate of 2% and a peripheral (but still AI-related) job rate of 7%.

A.2.1. AI-relevant skills from job postings data

To build the list of AI-relevant skills, we start by following Babina et al. (2024) to identify AI-relevant skills in the job postings data. We use job posting data from Lightcast (which bought Burning Glass—an early provider of high quality job postings data), which covers a near-universe of U.S. online job postings from 2010 through 2025. Each job posting includes a list of skills required by the employer, usually requesting 5 to 20 skills (the interquartile range is 5 to 18 skills, and the 5th and 95th percentiles are 1 and 34 skills, respectively).

There are about 32,000 required skills across all years with the skills standardized by Lightcast. Four of those standardized skills are direct and unambiguous markers of AI technology over the past decade or so: *Artificial Intelligence*, *Machine Learning*, *Computer Vision*, and *Natural Language Processing (NLP)*, which are often abbreviated as "AI", "ML", "CV", and "NLP", respectively. We refer to these as the "core AI" skills.

We then identify other skills closely related to AI by measuring how often each skill appears in the same job posting as the core AI skills across all years of job posting data (2010–2025). For each skill s in the Lightcast skill taxonomy, we define

$$w_s^{AI} = \frac{\text{\# of jobs requiring skill } s \text{ and (ML, NLP, CV, or AI in required skills)}}{\text{\# of jobs requiring skill } s} \tag{A.1}$$

where w_s^{AI} is the **AI-relatedness score** of skill s, which is the share of job postings requiring skill s that also require at least one core AI skill. We retain skills with high AI-relatedness score (with $w_s^{AI} \geq 0.6$) that also appear in at least 50 job postings (otherwise we pick up skills that only appear a few times: e.g., "Continuous-Time Markov Chain" appears in 1 job posting and "Wit.ai" in 17). The threshold of AI-relatedness of 0.6 balances the risks of excluding highly relevant skills and including less-AI-related skills. In practice, most AI-related jobs are identified by terms above the 0.9 level. This gives us a starting list of 335 AI-relevant skills. In Table A.6, we report the list of the top 50 skills with the highest AI-relatedness scores. Outside of the top four core AI skills, the three skills with 100% AI-relatedness scores are "Zero Shot Learning", "Rweka", and "AgentGPT". Of these, "Zero Shot Learning" is particularly relevant for foundation models and LLMs: it is an ML approach where an AI system performs a task without having seen labeled examples of that exact task during training. For example, models like ChatGPT can answer questions, summarize text, classify content, write code, translate, or reason about new prompts without being specifically trained on each individual task. "Rweka" is an R package that connects R to Weka, a Java-based collection of machine learning algorithms for data mining. Its strongest connection is to classical AI/ML. Finally, "AgentGPT" is a browser-based platform for creating and running autonomous AI agents. It is most relevant for agentic AI. It is encouraging to see that the methodology introduced in Babina et al. (2024) is able to pick up terms relevant for the recent AI waves such as generative AI (e.g., "AI Generated Code") or agentic AI (e.g., "Agentic Systems").

To avoid missing earlier AI skills that may have become less common over time, we

also add the skills from Babina et al. (2024)—based on an earlier vintage of the same underlying job postings data—that are missing in the starting list above. There are 12 such skills: Libsvm, Sentiment Classification, Torch (machine learning), Jung Framework, Latent Semantic Analysis, Stochastic Gradient Descent, Maximum Entropy Classifier, Expectation-Maximization (EM) algorithm, Classification Algorithms, Bayesian Networks, AI chatbot, ND4J (software).

To assess whether w_s^{AI} simply captures skills that "sound like AI" rather than skills that employers actually pair with AI requirements, we compare our co-occurrence-based measure to a semantic-similarity measure built from sentence embeddings. For each of the 335 seed skills, we compute the cosine similarity between the skill name and four AI anchor phrases ("artificial intelligence," "machine learning," "natural language processing," "computer vision") using a pretrained sentence embedding model (`all-MiniLM-L6-v2`).[52] The Pearson correlation between w_s^{AI} and this embedding-based score is only 0.06 (Spearman: 0.03), and similarly weak (0.04/0.01) when restricted to the 320 seed skills with an exact match in our final keyword dictionary. This weak correlation indicates that the two measures capture different things: terms like "Hugging Face Transformers" and "Seq2Seq" have high co-occurrence scores despite low lexical similarity to core AI terms, while "Text Mining" and "Machine Learning Algorithms" score higher on embedding similarity but co-occur less frequently with core AI skills because they also appear in generic analytics roles. This supports our use of co-occurrence rather than semantic similarity as the primary criterion for keyword selection.

A.2.2. AI keyword expansion and matching method

Having selected the seed AI skills in Section A.2.1, we next convert them into search terms that can be matched against job titles and descriptions in unstructured resume data from Revelio. A single AI skill can be written in many different variants in the text of jobs reported in resumes. For example, "Natural language processing" shows up as "NLP", "Natural Language Processing", "natural language processing", or as part of a longer phrase like "NLP engineer". "Large language model" shows up as "LLM", "LLMs" (plural), or as "large language model". To pick these up, we first expand each seed skill into the variants it plausibly takes in resume text. Specifically, we generate variants covering (i) the full skill name; (ii) standard abbreviations (when appropriate); (iii) common variants (singular/plural); (iv) words that are hyphenated (e.g., *multi-modal model*), separated (e.g., *multi modal model*) or merged together (e.g., *multimodal model*), and (v) shorter forms whose meaning is preserved (e.g., "*hugging face*" for "*Hugging Face Transformers*").[53]

Based on the structure of these keyword variants and our extensive experience with NLP and text matching, we define three matching methods that vary in how strictly each candidate keyword must appear in the text:

- **Method 1 (case-sensitive, word boundary).** The keyword must appear in the exact case

[52] We take the maximum cosine similarity across the four anchors as each skill's embedding-based AI-similarity score.

[53] In principle, we could have used an LLM to do "fuzzy matching" on these terms. We decided to go with a more traditional "bag of words" search approach because: 1) this method is very transparent and reproducible; and 2) the computation costs outweigh the benefits of a small increase in precision.

shown, with a word boundary on each side. A word boundary means that the keyword is delimited by the beginning and the end of the text string, whitespace, or punctuation. For example, "AI" matches the all-caps "AI" (the abbreviation for artificial intelligence), but not the lowercase "ai" or longer words containing that letter combination (e.g., "n**ai**cs" or "N**AI**CS").

- **Method 2 (case-insensitive, word boundary).** The keyword must appear with a word boundary on each side, but case does not matter. For example: "transformer" matches "Transformer" or "transformer", but not "Transformers" or "transformers".

- **Method 3 (case-insensitive, no word boundary).** The keyword need only appear somewhere in the text, ignoring case and word boundary. For example, a keyword "language model" picks up "language models" or "Multimodal Language Model" without us having to list these variants separately.

In this step, we transform each of the highly AI-relevant skills from the job postings into search terms across these three search methods. For example, "Machine Learning" gets searched as "ML" in Method 1 (case-sensitive, word boundary); as "ml/ai" and "ai/ml" in Method 2 (case-insensitive, word boundary); and as "machine learning" in Method 3 (case-insensitive, no word boundary). "Microsoft Cognitive Toolkit (CNTK)" gets searched for "CNTK" in Method 1 and "microsoft cognitive toolkit" in Method 3. "BERT (NLP Model)" gets searched for "BERT" in Method 1; "bert nlp" in Method 2; and "bidirectional encoder representations from transformers" in Method 3. This step expands the AI-relevant search term list to 433 unique terms.

To complement this manually created rule-based expansion, we use a large language model (`GPT-5-nano`) to suggest additional spellings and synonyms we might have missed. We feed to the model the AI-skill spreadsheet developed to this point one row at a time. For each row, the model sees the seed AI skill and its existing list of variants, and is asked to propose more variants. The prompt we give the model is the following:

```
We are trying to identify AI-related jobs in LinkedIn data by searching for
AI-relevant terms in people's job titles and job descriptions. For the
provided target term, suggest alternative ways people may write the same
concept on LinkedIn profiles or job postings.

Requirements:

- Only include highly related variants that improve recall.

- Include spelling variants, formatting variants, standard abbreviations,
official shorthand, branded sub-names, and very close synonyms only when they
are genuinely used as alternative search surfaces for the same concept.

- Do not include unrelated tools, competing tools, adjacent concepts, broader
categories, narrower subfields, job titles, or role labels unless they are
truly used as alternative names for the same concept.
```

```
- Do not include terms already covered by the existing terms list.

- If the existing terms already seem sufficient to capture the concept well,
return no additional suggestions.

- Prefer returning no suggestions over returning a borderline term.

- Flag terms that may create many false-positives.

For each suggestion, return: the term itself, its frequency in real text
(Very High / High / Medium / Low), its relevance to the seed concept (High /
Medium / Low), and its false-positive risk (Low / Moderate / High), along
with a short note.
```

For example, for the original skill in the job postings *Hugging Face Transformers*, the LLM suggests adding *HF transformers* and *HF transformer* (case sensitive and with word boundary). Lastly, we manually verify each GPT-suggested search term before adding it to the keyword list. For each candidate term, we conduct a web search and inspect the top results to assess whether the term is unambiguously AI-related. Terms whose results are clearly related to AI are retained, while terms yielding unrelated or ambiguous results are excluded. This step expands the AI-relevant search term list by an additional 89 unique terms (although by construction all of them are highly related to the already included terms).

A.2.3. False-positive screen of AI-relevant search terms

Next, we investigate whether we have included some AI-relevant search terms that also have a meaning not related to AI and, hence, identify some jobs as AI-related when, in fact, they are not (i.e., generating false-positive AI jobs). To do so, for every keyword, we use Revelio data from 2010–2024 to compute (i) the number of U.S. positions matched unconditionally by the keyword and (ii) the additional jobs identified by the keyword beyond those identified by the core AI-relevant terms, such as AI, ML, and NLP. The keywords with a similar number of jobs in groups (i) and (ii) are flagged as false-positive candidates. The idea is that if a term appears in job descriptions that do not also contain key AI terms, this potentially indicates that the term might be picking up jobs that are not AI-relevant. For example, the term "RL" (the abbreviation for reinforcement learning) appears in about 10,700 jobs, and 8,870 of those do not contain any core AI terms (we later learn that "RL" in resumes usually refers to "Ralph Lauren", an apparel company). We also flag a few terms as potentially high false-positive risk because they might be referring to something else like "feature engineer" (someone who works on features of a product, system, or platform not necessarily AI-related).

We identify 26 terms as potentially high false-positive risk. The flagged keywords are then run through two complementary checks: LLM classification at scale and human verification on a random subsample.

LLM classification at scale. For each suspect keyword, we collect **all** job positions from our public firm sample in Revelio whose title or description contains the keyword. We pass each such position, one at a time, to a large language model (GPT-5.4) with a structured prompt that includes the keyword that triggered the match, the position's title, and the position's description. The model is asked to judge whether the position should count as AI-related given the context. The instructions we give the model are the following:

```
You are reviewing LinkedIn profile records to determine whether each job
should count as AI-related under a broad research definition.

Count a job as AI-related (is_ai_related = 1) if the role develops, uses,
supports, manages, researches, deploys, sells, or materially relies on
AI-related skills or tools. The role does NOT need to be purely AI-focused.
If AI-related work is one meaningful part of the job, it counts.

Broad AI-related examples that should usually count:

- machine learning, deep learning, neural networks, LLMs, generative AI,
prompt engineering

- data science, applied science, quantitative modeling, predictive modeling,
statistical learning

- NLP, computer vision, speech, robotics, recommender systems,
search/ranking, AI optimization

- MLOps, AI infrastructure, model deployment, model monitoring, inference
systems

- AI product management, AI program management, AI solutions engineering, AI
consulting

- software, analytics, or data roles that explicitly use ML/AI models, AI
APIs, or related methods

Mark a job as NOT AI-related (is_ai_related = 0) only when it is clearly
unrelated to AI/data-science/ML work and the matched keyword is incidental,
ambiguous, or refers to something else.

Important rules:

- Use a broad definition: if the job uses or entails AI-related skills, count
it.

- If the description is missing, vague, or uninformative, infer from the
title and matched keyword and make your best guess.
```

```
- When evidence is mixed, lean toward 1 rather than 0.

- Do not require the role to build frontier models or have 'AI' in the
official title.

Return ONLY valid JSON with exactly these keys:

{"is_ai_related": 0 or 1,
 "reasoning": "short explanation, at most 15 words"}
```

We then aggregate the model's judgments across keyword-matched positions to estimate each flagged keyword's false-positive rate, defined as the share of positions matched by that keyword that are not actually AI-related. This gives us false-positive rates for all jobs identified as AI-related for the list of suspect search terms.

It is important to note that in the next step we validate the model's ability to identify AI-relevant jobs against the human classification by the authors on a random sample and find similar classification results.

Human verification on a random subsample. To verify whether the LLM correctly identifies AI- versus non-AI-relevant jobs in the above step in a systematic way, we draw a random subsample of keyword-matched Revelio positions for each suspect keyword, with 20 unique positions per keyword. We first ensure that this random sample is similar to the population of jobs with high potential false-positive risk: we use the same LLM prompt and compare the false-positive rates in this random sample to the universe of AI-related jobs identified in the step above. We confirm that the false-positive rates are very similar (the typical difference is less than 5%).

Next, using this random sample, each co-author independently reviews the position title and description and codes whether the position is genuinely AI-related, without seeing the LLM's classification. Aggregating these human labels gives a second, independent estimate of each suspect keyword's false-positive rate.

We then combine the LLM and the human false-positive estimates to decide whether to keep or drop each keyword. First, the agreement between the human and the LLM on which jobs are not AI-related and, hence, on false-positive rates, is quite high. Across the entire random sample, the human and the LLM classifications agree 91% of the time. This high agreement rate is reassuring, suggesting we can use this prompt to classify AI-tagged jobs as truly AI-related or as false positives.

Next, for each high potential false-positive term, we decide whether to keep it or drop it. If both estimates (the LLM's and the human's) of the false-positive rates are high (60% or higher), indicating that the keyword mostly matches non-AI positions, we drop it. For example, the candidate keyword *RL* (intended to capture the abbreviation for reinforcement learning) has a 95% false-positive rate by the LLM in the screened Revelio positions because it overwhelmingly referred to *Ralph Lauren* rather than *reinforcement learning*, and we therefore drop it. Another example—with a high false-positive rate of 90% that is dropped—is *loss adjusting*: in an ML context it refers to changing how a model's loss

function is calculated or weighted during training, but in resume data it typically refers to insurance context of investigating and assessing an insurance claim.

Next, if both false-positive estimates are low (less than 20%), meaning that both screens suggest the keyword mostly captures genuine AI jobs, we keep the keyword. For example, *language model* has only a 4% false-positive rate and we retain it. Similarly, we keep *feature engineer* with a 15% false-positive rate.

Finally, if the two estimates disagree or have moderate false-positive rates of under 40%, we manually review additional matches for that keyword case by case until there is enough evidence to decide. Overall, we keep 14 out of 26 suspect AI terms.

After these steps, our final AI-relevant keyword list contains 510 keywords (433 original terms from humans plus 89 from the LLM minus 12 terms with high false-positive rates).

A.2.4. AI technology subcategories

AI techniques have evolved across several waves. The big breakthrough occurred in the 2010s with the "Machine Learning Era" when models began learning directly from data rather than following predefined logic. During this time, models remained focused on narrow applications: they performed well in specific domains—such as playing games or detecting fraudulent transactions. By the 2020s, the technology had evolved into generative AI, enabling more general-purpose tasks such as drafting research reports. More recently, firms have begun experimenting with agentic AI: systems built on top of generative models that can plan, take actions across tools, and iteratively pursue multi-step goals with limited human prompting.

To capture different waves of AI, we further classify our AI-related keyword list into AI subcategories: early waves of AI labeled as "Wave I AI": machine learning, natural language processing, computer vision, and more recent ones labeled as "Wave II AI" (generative AI, agentic AI). We also construct a sixth category, general AI, that is about more general AI terms that are not specific to any other AI category (e.g., *artificial intelligence*, *AI ethics*). We do so by using LLM-assisted classification, followed by author review.

We pass the keywords to Claude 3 Opus and ask it to assign each keyword a probability of belonging to each of six categories. These are the five AI subcategories and a residual "general AI" category. The six probabilities sum to one, and the model also provides a short explanation for each classification. The full prompt is reproduced below:

```
Can you classify the attached AI technologies into one of five categories:
machine learning, natural language processing, computer vision, generative
AI, agentic AI? The categories are mutually exclusive. Create one column for
each category with the probability that technology belongs to the category.
All probabilities should sum up to one. Please also add a column with your
explanation for the classification. If a technology is general and does not
belong to a specific category, label the technology as "general AI" in a new
column. For each row, highlight the column with the highest probability.

Here are some examples in each category:

Machine learning: machine learning, deep learning
```

```
Natural language processing: NLP, chatbot

Computer vision: computer vision, object detection

Generative AI: GenAI, LLM, prompt engineering

Agentic AI: agentic system, CrewAI
```

Two co-authors independently review the output of the LLM classification and research ambiguous terms. We then assign each keyword to the AI category with the highest probability among the six if that maximum probability is equal to or above 0.5. Keywords for which the maximum probability is below 0.5 are placed in an "other" category. This produces a list of (mutually exclusive) terms specific to each of the six categories plus an "other" category. The examples for each category include: *ethical ai* for general AI, *pytorch* for ML, *doc2vec* for NLP, *rekognition* for CV, *ai generated code* for generative AI, and *crewai* for agentic AI. The "other" terms tend to be concepts and tools that are important across different AI categories: for example, "attention mechanism" has been important in early work with text in NLP applications, but is also used in generative AI models. Of course, many of our AI search terms span across interrelated AI waves as AI progress tends to be cumulative with newer AI waves utilizing new and existing AI-related technologies, tools, and concepts. For example, while *deep learning* (classified as the ML subcategory with 85% probability and hence belonging to that subcategory) has emerged as an important methodology within a broader machine learning family in the 2010s, it is also an "engine" underlying many generative AI applications. So this subclassification should be thought of as approximate.

A.2.5. Identifying AI jobs in worker resume data

Using our comprehensive AI-related keyword set described above, we apply it to the worker resume data from Revelio Labs to search for AI-related jobs in our U.S. public firm sample. Revelio provides standardized data on each job a worker reports in their online resume with the job's starting and ending dates, and, critically, the job title and its description (we use both of them to search for our AI-related terms). While Revelio offers international coverage of worker resumes, we focus on U.S.-based positions for identifying AI-related jobs since, as we report in Part A.1.7, the U.S. coverage of employees of U.S. public firms tends to be more complete than its international coverage and since our keyword list is in English (creating and validating similar lists of AI-relevant search terms for all other languages using the steps above would be a major undertaking, which we leave for future research).

We first construct our main firm-level AI measure that spans all AI-related keyword terms. For each full-time U.S. job, we search its job title and job description against our comprehensive AI keyword list, applying the keyword-specific matching rule defined in

Section A.2.2.[54] A job is classified as AI-related if at least one AI keyword is matched in either the job title or the job description. We also construct firm-level AI measures for the AI subcategories described above: general AI, machine learning, natural language processing, computer vision, generative AI, agentic AI, and "other". We repeat the search for AI-relevant jobs but now use the AI-related keyword list specific to each AI subcategory. In addition, we construct firm-level AI measures that intend to capture early (2010s) versus more recent (2020s) waves of AI. We define "Wave I AI" jobs as jobs matching at least one of the ML, NLP, or CV search terms, and "Wave II AI" jobs as those that match at least one of the Gen AI or Agentic AI search terms. A job can be in several subcategories if it contains keywords from different AI subcategory lists.

To construct firm-level measures of AI-skilled jobs, we aggregate the firm-worker-job-year panel to the firm-year level by calculating the share of a firm's jobs that are AI-relevant in each year. Specifically, let $Job_{f,t,\text{US}}$ denote the set of full-time U.S. positions at firm f in year t, and let $AI_{f,t,\text{US}} \subseteq Job_{f,t,\text{US}}$ denote the subset classified as AI jobs. The firm's AI job share is defined as

$$\text{Share AI Jobs}_{f,t} = \frac{AI_{f,t,\text{US}}}{Job_{f,t,\text{US}}}.$$

We repeat this exercise by using the counts of AI-relevant jobs specific to each AI subcategory, and calculate the shares of AI workers for each subcategory (the number of workers in an AI subcategory divided by the total U.S. employment). We also calculate firm-level AI shares for "Wave I AI" jobs (NLP, ML, or CV) and for "Wave II AI" jobs (Gen AI or Agentic AI). Note that the sum of AI shares across subcategories does not have to add to the main AI share (generated across all search terms) since a job can be classified into several AI subcategories if it has search terms from several AI subcategories. To avoid this double counting and create mutually exclusive job categories, we also generate three additional versions of AI jobs: those *only* containing search terms from the "Wave I AI" search term list, those *only* containing search terms from the "Wave II AI" search term list, and those that contain search terms from *both* lists.

A.2.6. AI-job false-positive rate

We next offer several analyses of our new data on AI-related jobs. All of these statistics are based on the sample of U.S.-based workers of U.S. public firms matched to Revelio that have employment coverage between 0.05 and 3.

We start by examining the classification error of AI-related jobs: jobs that our methodology tags as AI-related but might not actually be AI-related, either because a search term might be used in non-AI contexts or because people put AI-related terms in their resumes without doing any significant development or implementation of AI in their jobs.[55] We also examine this error across all firms and by firm size (Figure C.4 shows that the share of AI-related jobs is disproportionately concentrated in larger firms and the time-series of

[54] En-dash and em-dash characters in job titles and descriptions are normalized to hyphens, and consecutive whitespace is collapsed to a single space. We also exclude positions with titles containing terms that indicate internships, part-time work, student roles, postdoctoral roles, or other temporary positions.

[55] Part A.2.3 describes how we remove AI terms with high false-positive rates. While this procedure reduces false-positive AI matches, it is unlikely to eliminate all false-positive AI matches.

these shares are more volatile for smaller firms).

We start with Table A.7 which reports false-positive rates by firm-size terciles. Using the same LLM prompt and model as in Section A.2.3, we audit *all* 5,084 AI-classified job-year observations in the small-firm tercile, and equally sized random samples from the middle- and large-firm terciles for symmetric comparison (we draw random samples because of the large number of jobs in these larger-firm terciles). The false-positive rate declines monotonically with firm size, from 5.47% among small firms to 3.38% among middle-tercile firms and 2.03% among large firms. Keyword-based AI classifications are therefore somewhat noisier among smaller firms, but the estimated false-positive rate remains modest across all three groups. These false-positive rates translate into an overall moderate false-positive rate of 2.13% across all firm sizes (by weighting the false-positive rates for each firm size tercile by the share of jobs in small, medium, and large firms, specifically 2.13% =0.55%*5.47%+5.79%*3.38%+93.66%*2.03%).

We next repeat the false-positive analysis for each AI subcategory by randomly drawing 5,000 AI-related tagged jobs in each AI subcategory. Table A.8 shows that the NLP category has the highest false-positive rate of 2.7%: it turns out that the NLP-specific term "BERT" (Google's early AI model) can also stand for "Bit Error Rate Tester", which is a test instrument used to measure how accurately a digital communication link transmits data and is not related to AI, pushing the false-positive rate up for the NLP subcategory. The general AI category has the second highest false-positive rate of 1.8%. Other subcategories have the following false-positive rates: 0.7% for agentic AI, 1.2% for CV, 0.5% for gen AI, 1.6% for ML, and 0.4% for the "other" category. Overall, all these false-positive rates are small.

A.2.7. Summary statistics on AI-related jobs

We next provide summary statistics on the AI search terms, AI-related jobs, and firms with highest AI shares and job counts. As in the previous Part A.2.6, all statistics below are based on the sample of U.S.-based workers of U.S. public firms that have the employment coverage between 0.05 and 3 (unless explicitly stated otherwise).

We start by plotting the number and the share of jobs that are AI-related in Figure 1 panels (a) and (b), respectively. We see a strong upward trend in both figures, suggesting there are more AI-related jobs over time overall and as a share of all jobs. By the end of the sample in 2024, there are more than 300,000 AI-related jobs, corresponding to about 2.1% of all jobs. These figures represent aggregate trends (the number of AI jobs is the sum over all AI jobs each year, and the share of AI jobs is the ratio of all AI jobs to total jobs). To examine firm-level AI workforce dynamics, we recalculate AI worker shares as the average of AI shares across all firms in a year and AI worker numbers as the average number of AI workers per firm. We plot these averages in Figure A.5. At the end of the sample, on average, a U.S. public firm has 78 AI-related jobs (panel (a)) and 1.5% share of AI-related jobs (panel (b)). Figure A.6 shows that the distribution of these AI-related job numbers and shares is quite wide across firms: this figure re-plots the means from Figure A.5 and also adds colored bands for different parts of the distribution: 90–95th percentile in the darkest blue, 75–90th percentile in the medium-light blue, and 50–75th percentile in the lightest blue. While the median firm has very few AI jobs, Figure A.6 shows that a firm at the 95th percentile has 213 AI jobs and 6.7% AI share.

Figure 5 shows aggregate AI shares by NAICS-2 industry sector at the beginning (2010), the middle (2018), and the end of our sample (2024). It shows that the tech sectors have the highest shares of AI workers, although the growth in AI-related workers over time occurs across all sectors.

We also examine aggregate AI shares over time by AI subcategories. Figure 3 shows that the growth in AI job shares in "Wave I" AI technologies (ML, NLP, CV) has either slowed down (ML) or started to decelerate (CV, NLP) at the end of the sample (although the levels are still high, especially for ML, which identifies about 0.9% of jobs as AI-related in 2024). Instead, more recent waves of AI (generative AI and agentic AI) accelerate very rapidly at the end of the sample, with a 0.45% share of generative AI jobs in 2024. Table A.9 tabulates the numbers underlying these plots from Figure 3. Figure 4 shows these trends even more clearly by aggregating ML, NLP, and CV into a single category "Wave I" AI, and generative AI with agentic AI into "Wave II" AI. In relative terms, "Wave I" AI jobs are about 1% and "Wave II" AI jobs are about 0.5% of jobs in 2024. Note that the sum of these two shares is lower than the aggregate 2.1% share because these AI categories do not rely on search terms from "General AI" category (e.g., "AI") or "Other" category (e.g., "operationalizing ai").

We continue by examining which AI-relevant keywords identify most AI jobs overall and by six AI subcategories in 2024 (the last year of our resume sample). Table A.10 reports the top 50 keywords with the most AI jobs. The top 5 terms are *AI, machine learning, ML, generative ai, ai/ml*. Table A.11 shows the top 10 search terms for each specific AI subcategory and all search terms make sense to us. Besides the core AI terms, the top three terms by job counts are as follows: for general AI, it is *ai research, ai innovation, ai security*; for ML it is *deep learning, numpy, tensorflow*; for NLP—*conversational ai, sentiment analysis, BERT*; for CV—*object detection, image recognition, opencv*; for generative AI—*prompt engineering, chatgpt, retrieval-augmented generation*; and for agentic AI—*langchain, ai agent, langgraph*. We also examine which words most often appear in job titles and descriptions for each AI subcategory: Figure 2 shows that while each AI subcategory has its own specific AI words, one common and important word is *data*: this makes sense since any type of AI needs data to be useful.

We provide several examples of AI-related jobs for each AI subcategory in Table A.12. In the general AI category, we tend to observe leadership jobs, jobs that leverage AI but where AI is not the main focus, and cross-functional roles that help build AI into products and processes. In more specific AI subcategories (ML, NLP, CV, generative AI, agentic AI), AI-related roles tend to be more technical and data heavy. Agentic AI-related jobs tend to contain generative AI related terms: this is not surprising since agentic AI relies heavily on GenAI capabilities.

We next examine more systematically job titles of AI-related jobs. We use job titles self-reported by workers, which tend to be less standardized than occupations: for a functionally similar job, different firms might use different job titles.[56] Table A.13 shows

[56]We lightly standardize job titles by aggregating across closely related title variants: seniority and level modifiers, such as "Senior," "Staff," "Principal," "Lead," "II," and "III," are combined when the underlying role is the same.

30 job titles with the most AI-related jobs in 2024. Intuitively, it shows that AI-related jobs tend to be in software, engineering, scientific, and data-related positions. Table A.14 shows the top 10 job titles (by the number of AI jobs) for each AI subcategory. We observe a remarkable overlap of job titles across all AI subcategories.

Using jobs' occupation data standardized by Revelio, Table A.15 shows the top 30 occupations with the highest *shares* of AI workers in 2024. The top three occupations are: 1) "Computer and Information Research Scientists" with a whopping 0.66 AI share, 2) "Data Scientists" (0.43), and 3) "Mathematical Science Occupations, All Other" (0.39). Table A.16 reports top 30 occupations with the highest *number* of AI jobs in 2024: the top three are 1) "Data Scientists" with 50,403 AI jobs (with 0.43 AI share); 2) "Software Developers" with 40,418 jobs (0.06 AI share), and 3) "Computer and Information Research Scientists" with 23,985 jobs (0.66 AI share). Upon inspection, these top AI-related occupations intuitively cover many engineering, data, and scientific occupations, but also domain area experts (e.g., marketing managers, designers, financial quantitative analysts) whose knowledge is needed to apply AI effectively to solve business problems. These tables also highlight that using occupations to identify AI-related workers is not a good strategy since even the very top ones have high shares of non-AI-related workers.

We next examine education, compensation, and job seniority of AI- and non-AI-related jobs (we do not examine these characteristics at the worker-level since many workers start in non-AI-related roles and then transition into AI-related roles). We find that AI-related workers are more educated and more highly compensated, and tend to be in more senior roles. Among workers who report college education, Figure A.7 shows the shares of workers with the highest educational attainment being a Bachelor's, a Master's, or a Doctoral degree for AI-related jobs (left bar) and non-AI-related jobs (right bar). Perhaps unsurprisingly, AI-related workers tend to be more educated: 11% of AI-related workers have a Doctoral degree (versus 4% for non-AI workers), 52% have a Master's degree (34% for non-AI workers), and 37% have an undergraduate degree (62% for non-AI workers). We also examine the mean differences in the total compensation of AI and non-AI-related jobs (the compensation data is imputed by Revelio by using industry, location, and occupation of each job, so it is inherently noisy). Figure A.9 shows that, on average, AI-related workers have 62% higher compensation than non-AI-related workers: the average total compensation is about $224 thousand for AI-related and $138 thousand for non-AI-related jobs.

Lastly, we show that AI-related workers tend to be in more senior roles than non-AI-related workers in Figure A.8.[57] AI-related jobs are much less concentrated in junior roles than non-AI jobs: 20% of AI-related positions are junior, compared with 53% of non-AI positions. Instead, AI-related positions are more concentrated in mid-level and senior roles. Mid-level positions account for 42% of AI jobs, compared with 29% of non-AI jobs, while senior positions account for 38% of AI jobs, compared with 18% of non-AI jobs.

We switch gears to examine which specific firms have large AI-related workforces.

[57] Revelio assigns jobs to seven seniority levels: 1 corresponds to entry-level roles, 2 to junior roles, 3 to associate roles, 4 to manager roles, 5 to director roles, 6 to executive roles, and 7 to senior executive roles. We aggregate these seven levels into three broader groups: junior roles include levels 1–2, mid-level roles include levels 3–4, and senior roles include levels 5–7.

It is worth a reminder that these firm-specific AI shares of workers are approximate for the reasons described in Part A.1.7, and should be taken as a proxy for how AI-focused firms are. For firm-specific statistics, we keep firms with at least 50 jobs in Revelio and 50 employees in Compustat to avoid reporting on very small firms. We start by plotting the shares of AI workers over time for each firm with the highest AI share in 2024 within each NAICS-2 industry sector. Firms' industry codes are from Compustat. To observe time-series trends, we require a firm to have at least 5 years of data in Figure A.11 and 10 years in Figure A.12. While most firms across the two figures show an upward trend in their AI shares, we keep both figures because they show different sample composition patterns. Starting with Figure A.12 which includes firms with longer time-series, we see some of the well-known AI-intensive incumbents like "Expedia", "Visa", and "Nvidia".[58] Turning to Figure A.11 which includes firms with shorter time series, we see that many firms with the highest AI shares in their respective industry are newly public firms that are often "AI native". A few examples follow. "Corteva" in Agriculture focuses on seed, crop protection, and digital agriculture: AI helps predict pest-protein structures, identify crop-protection molecules, model thousands of molecule candidates, accelerate biological discovery, optimize fermentation, guide field decisions such as fungicide timing, and support regulatory dossier preparation. "DraftKings" in Arts/Entertainment (digital sports entertainment and gaming firm) uses AI in live-betting pricing and technology platform. "Upstart" in Finance and Insurance reports that its AI models are central to its lending marketplace and are used for credit-risk assessment, income and identity fraud detection, borrower acquisition, loan stacking, default and prepayment prediction, and servicing. "Ontrak" (in Health Care) uses AI-enabled predictive algorithms and claims analytics to identify members with behavioral-health needs and medical-cost risk. "C3.AI" (Information) is an enterprise AI software company selling AI platforms, applications, and generative-AI products.

Next we report top 3 firms with the highest *shares* and *numbers* of AI workers in 2024 within each NAICS-2 industry sector in Tables A.17 and A.18, respectively.[59] The two tables highlight a distinction between AI intensity and AI scale: the AI share ranking tends to identify firms for which AI-skilled labor is central to the organization, while the AI job count ranking captures firms with the largest absolute deployment of AI workers.

In Table A.17, the intensity leaders are typically AI-native, data-platform, or highly specialized firms. "Innodata" has the highest AI share in Professional and Business Services, with 71.7% of its U.S. workforce in AI-related roles. "C3.AI" leads Information with a 52% AI share, "Coupang" leads Retail Trade with 21%, "Nerdy" leads Education Services with 18%, "Tempus AI" leads Health Care with 11%, and "Upstart" leads Finance and Insurance with 10%. Many of these firms have modest AI headcounts, but AI accounts for a large fraction of their workforce. They also tend to be relatively small firms by overall headcount.

[58] In agriculture, Figure A.12 shows zero AI shares for "Alico" (a US-based agribusiness and land management firm headquartered in Fort Myers, Florida): none of the firms in this sector with at least 10 years of data have a positive AI workforce in 2024.

[59] We do not include firms in agriculture since there is only one firm that has non-zero AI workers and we already reported on it in Figure A.11.

The scale leaders in Table A.18, by contrast, are large incumbents with many U.S. workers. Unsurprisingly, the Information sector has the firms with the highest numbers of AI jobs: "Microsoft", "Alphabet", and "Meta" lead the Information sector with 26,169, 21,385, and 15,075 jobs, respectively. "Amazon" and "Walmart" lead Retail Trade with 14,668 and 1,876 AI jobs, respectively. In Finance and Insurance, "JPMorgan Chase", "Capital One", and "Wells Fargo" lead with 4,527, 3,284, and 2,699 AI jobs, respectively. In Health Care, "CVS Health" leads with 2,173 AI jobs. In most sectors, the AI job share and the AI job count leaders differ, underscoring that AI intensity and AI scale capture distinct subpopulations of firms.

Because technology firms are natural AI adopters and dominate many aggregate AI rankings, we next separate technology and non-technology firms to examine top 30 firms with the highest numbers and shares of AI workers across these two groups of firms. Within the tech sectors (Information; Professional and Business Services), Tables A.19 and A.20 show that AI jobs are concentrated among large incumbent platforms and enterprise-technology firms, while AI intensity is concentrated among specialized AI and data firms. "Microsoft", "Alphabet", "Meta", and "IBM" together account for about 72,000 AI jobs in 2024. The next tier includes "Accenture", "Salesforce", "Oracle", "Adobe", "Verizon", "AT&T", "Booz Allen Hamilton", "ServiceNow", and "Intuit", each with more than 1,900 AI jobs. The AI share ranking, however, is led by smaller and more specialized AI and data-intensive firms. "Innodata", "C3.AI", "Coveo Solutions", "Baidu", "Cerence", "Veritone", "SoundHound AI", "LivePerson", and "Rekor Systems" report AI shares ranging from about 20% to 72%.[60] Thus, within technology, large platforms dominate the absolute stock of AI workers, while specialized AI and data-intensive firms are the most AI-intensive.

Outside the tech sectors, Tables A.21 and A.22 suggest a similar distinction across a broader set of industries. The non-tech AI share ranking is led by firms in semiconductors, AI hardware, biotechnology, fintech, education technology, and data-intensive platforms. Given the nature of the firm, "Nvidia" has one of the highest non-tech AI shares and one of the largest non-tech AI workforces, with an AI share of 30% and 8,965 AI jobs. Other high-share non-tech firms include "Xpeng" at 36%, "Absci" at 23%, "Coupang" at 21%, "Nerdy" at 18%, "Ambarella" at 18%, "Weride" at 18%, and "Ceva" at 16%. These firms illustrate that high AI intensity outside technology is concentrated among smaller firms whose products, production processes, or platforms are more closely tied to AI. The non-tech count ranking is instead led by large incumbents. "Amazon" has 14,668 AI jobs, followed by "Nvidia" with 8,965, "Apple" with 7,900, "Intel" with 5,430, "JPMorgan Chase" with 4,527, "Cisco" with 3,972, "Capital One" with 3,284, "Dell" with 3,163, "Wells Fargo" with 2,699, and "Tesla" with 2,545.

Taken together, these tables show that AI-skilled labor has spread well beyond a narrow set of technology firms, but along two different margins. The firms with the highest AI shares are often smaller AI-native or data-heavy companies, where AI is central to the business model. The firms with the largest AI workforces, by contrast, are mostly large public firms in technology, manufacturing, finance, retail, and health care.

It is interesting to contrast these data points (showing that the most AI-intensive firms

[60] Baidu has only a limited presence in the U.S.

tend to be smaller firms) with the AI intensity across all firms by firm size. Figure C.4 shows that by firm size terciles (small, medium, large), the shares of AI workers are largest for big firms, followed by medium-sized firms, and smallest for small firms. This is due to lower average AI adoption rates by small and medium-sized firms than by large firms: in 2024, the share of firms with at least one AI job was 28% in the small-firm tercile, 73% in the medium-firm tercile, and 94% in the big-firm tercile.

A.2.8. Comparison to firm-level data from Cognism

We next compare our new firm-level data built from resume data from Revelio Labs to the similar data we have built from resume data from Cognism (which we introduced in Babina et al. (2024)). Since the underlying source of resume data has changed and since we updated the methodology for identifying AI-related jobs, we want to investigate how employment count as well as AI measures compare across the two resume datasets.

Table A.23 reports the correlations between Revelio and Cognism AI measures in each year from 2010 to 2021 (last year when Cognism data were available) for the regression sample. The correlations reported in the table are generally high, indicating that the Revelio and Cognism datasets paint a broadly consistent picture of the firms' workforce while not being identical. Agreement is strongest for total job counts, where correlations are uniformly high—ranging from 87% in 2010 to 94% in 2021—and rise steadily over the period, suggesting the two sources track overall employment levels closely. The datasets are also positively correlated on the AI share, with agreement generally strengthening over time: going from 45% in 2010 to 72% in 2021. The consistent positive correlation indicates that the two sources move together in capturing the AI intensity of hiring.

A.2.9. Comparison to firm-level data from Lightcast

Because some resumes are updated retrospectively, workers might relabel past positions using AI terminology that became fashionable only later (Bloom et al., 2026). Job postings data provide a natural check for this issue. They are contemporaneous employer records, so if resume-reported AI hiring reflects genuine firm demand rather than relabeling, the years in which a firm's workers report starting AI positions should coincide with the years in which the firm advertises AI-related vacancies. For each firm-year over 2010–2024, we compare the number of new AI hires in the Revelio resume data with the number of the firm's Lightcast job postings that list the same AI-related skills from our keyword list. We also restrict the measure to skills that (i) are counted only from their first appearance in the postings data, so late-arriving vocabulary cannot backdate resume positions, and (ii) appear in both the firm's job postings and its new hires' resumes within the firm-year.

Appendix Figure A.13 shows a binned scatter plot of new AI resume jobs against AI-related postings for each firm-year, residualized on year fixed effects, total postings, and total U.S. resume jobs. The partial correlation is 0.89, and 0.74 when excluding firm-years with no AI activity in either source.

A.3. Construction of AI Job Categories

In our next step, we investigate what kinds of jobs employees do in our AI-related job sample. For this, we draw a random sample of 300 AI jobs each year from 2010 through 2024, and categorize this sample of 4,500 jobs into several functional roles. As a first step,

we read through a random sample of thousands of job descriptions to build intuition about the most common types of AI-related jobs.

Two-step classification design. We implement the classification in two stages. In *Step 1*, the LLM constructs a mutually exclusive and collectively exhaustive taxonomy of 8–12 functional categories from the job sample. In *Step 2*, we freeze this taxonomy and assign each job to exactly one category based on the worker's job title and job description. For the primary results reported in Tables A.24 and A.25, we use *GPT-5.4* with a "low" reasoning effort to run Step 1 and Step 2 on all 4,500 jobs, yielding the 11-category taxonomy in Table A.24 and the category counts in Table A.25.

Initial pipeline and alternative approaches. We also evaluated an embedding-based clustering approach as a lower-cost alternative to LLM classification. The resulting clusters overlapped only partially with the LLM taxonomy, and boundary cases proved difficult to interpret consistently, so we did not adopt this method. Turning to the stability of the LLM approach itself, we find that independent Step 1 runs with a fixed prompt and fixed model settings yield taxonomies with the same number of categories and broadly similar structure. We therefore fix the primary taxonomy from the initial classification run and proceed to validating Step 2 reproducibility and classification accuracy.

Step 2 reproducibility. The following checks are separate from the primary classification run above. We assess Step 2 stability using two frozen taxonomies: the primary taxonomy (11 categories in Table A.24) and an alternative 12-category taxonomy from a separate Step 1 run with *GPT-5.6* (high reasoning). For each taxonomy, we run Step 2 twice independently on the same 4,500 jobs, using *GPT-5.4* with temperature set to zero and reasoning effort set to none. The two passes agree on 4,326 of 4,500 jobs (96.1%) under the primary taxonomy and on 4,335 of 4,500 jobs (96.3%) under the *GPT-5.6* taxonomy. Step 2 is therefore highly reproducible under both codebooks.

Human audit. We further audit classification quality on a random subsample of 200 jobs from the same 4,500 jobs. Blind to the LLM labels, we manually assign each job to one category, once using the primary codebook and once using the *GPT-5.6* codebook. The manual labels agree with the first Step 2 validation pass on 176 of 200 jobs (88.0%) under the primary taxonomy and on 178 of 200 jobs (89.0%) under the *GPT-5.6* taxonomy. These checks indicate that LLM classifications align closely with human judgment under both codebooks.

```
AI Job Classification Task
```

You are an expert labor economist classifying LinkedIn resume entries. Each row in the attached CSV or Excel file represents one job listed on a worker's LinkedIn profile. title_raw is the worker's self-entered job title. description is the worker's own description of their responsibilities in that position. It is not an employer job advertisement. These jobs were selected because they contain AI-related keywords in either the title or description. However, not all of them represent workers whose primary responsibilities involve AI development or implementation. Many only mention AI tools or AI-related terminology. Your task is to distinguish these different types of AI involvement.

Classification Objective

Classify each job according to the worker's primary role in the AI production, deployment, governance, or use process. The goal is not to determine whether a worker "works in AI," but rather how they contribute to AI. The categories should distinguish between workers who: build AI systems, support AI development, deploy or integrate AI, manage AI initiatives or coordinate with AI workforce on AI implementation, govern AI, claim implementation of AI tools in work processes but work in non-AI related roles, use AI tools in non-AI work, or merely mention AI without meaningful involvement.

Step 1: Construct a Taxonomy

Before classifying the jobs, develop a taxonomy consisting of 8-12 mutually exclusive and collectively exhaustive (MECE) categories. The taxonomy should be: comprehensive enough to cover all jobs, internally consistent, applicable across all uploaded files, based on the actual distribution of jobs in the data, stable so that similar jobs receive the same category. The categories should reflect the worker's function within the AI ecosystem, rather than industry, employer, seniority, or job title.

The following categories are intended as starting points and may be refined or reorganized if doing so improves consistency:

Direct AI model development, training, or engineering

Data engineering, data science, or annotation supporting AI model development

AI implementation, deployment, integration, or MLOps

AI product, project, or cross-functional program management and coordination with AI workforce on AI implementation

AI governance, ethics, compliance, documentation, education, or internal enablement

Business leaders or managers who oversee AI development and implementation without directly building models and who work in AI related roles

Business leaders or managers who oversee AI adoption without directly building models and work in non-AI related roles

Workers who use AI tools in their work but are not responsible for AI development or deployment

Incidental AI mention (AI keywords appear, but there is insufficient evidence that AI is a meaningful part of the job)

Step 2: Classify Every Job

Assign exactly one category to every row. The category should reflect the worker's primary responsibility, not every activity they mention. If multiple AI-related activities appear, classify according to the activity that appears most central to the role.

Evidence Hierarchy

Use the following evidence in order of priority:

Job description (highest priority)

Job title

AI-related keywords

If the description conflicts with the title, trust the description. Do not infer responsibilities that are not supported by the text.

Classification Principles

Judge the work actually performed.

Do not combine jobs with incidental AI mentions and those with insufficient evidence into one category. If there is insufficient evidence, try your best to assign it to a category; if you really cannot make an informed decision, then put it in its own ‘‘insufficient evidence’’ category.

Ignore seniority when assigning categories.

Ignore company reputation.

Ignore formatting artifacts, HTML, encoding issues, or mojibake.

```
Do not assume technical AI expertise simply because AI tools or AI
terminology are mentioned.

Mentioning ChatGPT, Copilot, LLMs, or similar tools alone does not imply that
the worker develops or deploys AI systems.

Distinguish between:

creating AI systems,

supporting AI development,

deploying AI,

governing AI,

managing AI initiatives,

adopting AI tools within organization,

using AI as a productivity tool,

and merely mentioning AI.

If evidence is insufficient to distinguish between two categories, choose the
broader, less specific category rather than inferring technical AI
responsibilities.

If job description is missing or short, rely on job titles in the following
way: if job description or title has highly AI specific terms (e.g., ‘‘ML
engineer’’, ‘‘gen AI developer’’, ‘‘Deep learning researcher’’) then classify them
into one of the highly AI-relevant categories that is closest by nature of
the job title. Do not put these jobs into ‘‘insufficient evidence’’ jobs.
```

Table A.24 lists the functional categories and their definitions. The taxonomy covers a broad set of AI functions within a firm, while separating substantive roles in AI development, deployment, management, and governance from jobs with incidental AI mentions. Table A.25 reports the distribution of the 4,500 AI jobs across these categories. Only 336 jobs (7.5%) are classified as incidental or unsubstantiated AI mentions: after examining these jobs, they often contain legitimate mentions of AI-related keywords and some AI-related functions but are not focused exclusively on AI implementation within firms, or do not provide enough detail in job descriptions to classify them into one of the categories with high-enough confidence.

Most of the AI-related jobs are heavily involved in internal AI development, management, and support of firm AI efforts. About 45% of jobs are in core AI development: 650 jobs (14.4%) are classified as AI research and model development; 469 jobs (10.4%) as AI

application and systems engineering; 273 jobs (6.1%) as AI platform, infrastructure, and MLOps; 531 jobs (11.8%) as domain analytics and predictive modeling; 51 jobs (1.1%) as AI data, annotation and evaluation support, and 1.4% of jobs provide governance of AI development. 22.7% of jobs are managing AI efforts: AI business leadership and functional oversight jobs (10.9% of AI jobs) provide executive, business-unit, or functional leadership over AI-related teams, portfolios, or investments; and AI product, program and delivery management jobs (11.8%) lead AI product strategy, program execution, or cross-functional delivery for AI products and features. A significant share of jobs involve AI go-to-market, solutions and advisory (7.4%) that help sell and support AI products and services.

Finally, 17% of the jobs are general technical workers who do not seem to exclusively work on AI development and implementation but claim using AI technology in their jobs (e.g., "Won company-wide engineering contest classifying websites using *machine learning*"; "I am a Software architect with expertise in data science, cloud, *machine learning*, and cybersecurity leveraging languages C, C++, Python, Java"; "Develop new business ideas in the Augmented Reality, Virtual Reality and *Artificial Intelligence* spaces").

These patterns suggest that our resume-based AI measure primarily captures jobs that contribute to firms' internal AI development and deployment as well as AI product management, support and monetization, rather than merely workers using external AI tools. When we force a model to have a separate classification category for *"Non-AI role using AI tools as productivity aid"*, the model classifies only 0.9% of jobs into this category.

We also examine how these AI job categories are distributed within the tech and the non-tech sector firms. In Table A.26, we show that the share of AI jobs in AI management roles ("Management") is about the same in the two sub-samples. However, non-tech sector firms have a substantially higher share of AI jobs in "Core AI development" (48% vs. 43% in tech firms), and also higher shares in "Other AI" jobs (27% vs. 22% in tech), while having a much smaller share of AI workers in selling and supporting AI products and services (3% vs. 12% in tech). This is consistent with tech firms being more likely to use AI to build products and services to sell to others while non-tech firms invest in AI to build in-house AI capabilities.

A.4. Measures of Productivity

This Appendix provides definitions of our two firm-level TFP measures. Our baseline measure is revenue TFP (Appendix A.4.1), and an alternative measure based on the procedure of Imrohoroğlu and Tüzel (2014) is used in robustness checks (Appendix A.4.2).

A.4.1. Revenue TFP

Our revenue TFP is the residual from an industry-level production function estimated by OLS. For firm i in industry s and year t, we estimate the following specification separately for each industry,

$$\log y_{it} = \mu_i + \mu_t + \alpha_s^l \log(l_{it}) + \alpha_s^k \log(k_{it-1}) + \varepsilon_{it}, \tag{A.2}$$

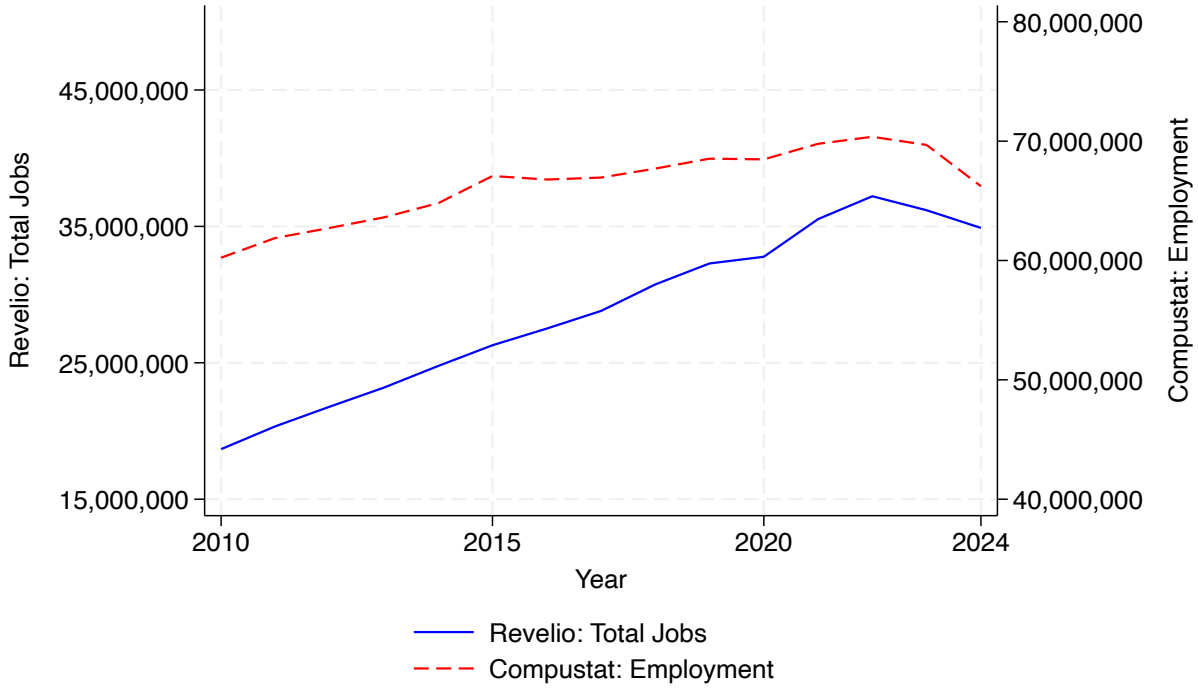
where y_{it} is real sales, l_{it} is employment, k_{it-1} is the beginning-of-period capital stock, and μ_i and μ_t are firm and year fixed effects. The capital stock is constructed using the

perpetual inventory method. Revenue TFP is the estimated residual $\hat{\varepsilon}_{it}$. This measure is specific to Cobb–Douglas production functions.

A.4.2. TFP based on the control-function approach

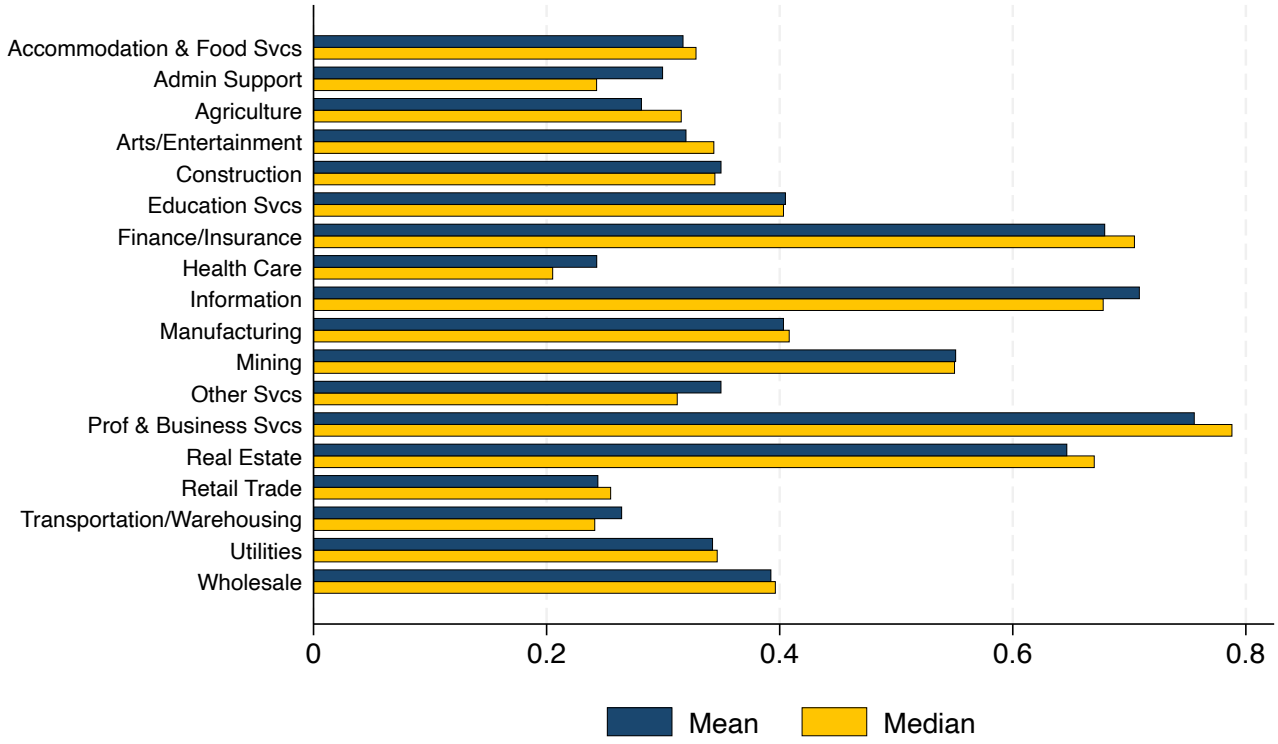
We follow the implementation of Imrohoroğlu and Tüzel (2014) using the control-function approach by Olley and Pakes (1996) to construct an alternative measure of firm-level TFP. This approach addresses the correlation between firms' input choices and unobserved productivity. First, we measure each firm's real value added (sales less materials), labor input (number of employees), and capital stock from Compustat, with output and inputs deflated by industry and aggregate price indices. The capital stock is measured from gross property, plant, and equipment and deflated using investment prices from the years in which the capital was purchased. Second, we estimate a value-added production function in labor and capital separately for each industry, using the procedure of Olley and Pakes (1996), which uses firms' investment to control for the correlation between unobserved productivity and input choices. Third, firm-year TFP is the residual from this production function, equal to log value added minus the estimated contributions of labor and capital. To expand the sample size, we lower the minimum employment requirement from 100 to 50 employees, matching our baseline sample; we relax the requirement that a firm have at least two years of Compustat history; and rather than drop industries with fewer than five firms, we pool them into their 2-digit industry when the pooled industry exceeds five firms. Because the procedure imposes many sample restrictions and excludes industries where physical inputs are not well defined, the measure is available for a smaller subset of firms. For instance, it drops firms in the finance and utilities sectors, firms with missing investment, and firms with low sales or material costs. Our resulting productivity measure has a correlation of 0.986 with the productivity data in Imrohoroğlu and Tüzel (2014). We use it only in robustness checks, as a productivity measure that is methodologically independent of our revenue-TFP residual.

Figure A.1. Revelio and Compustat Employment for Matched Public Firms



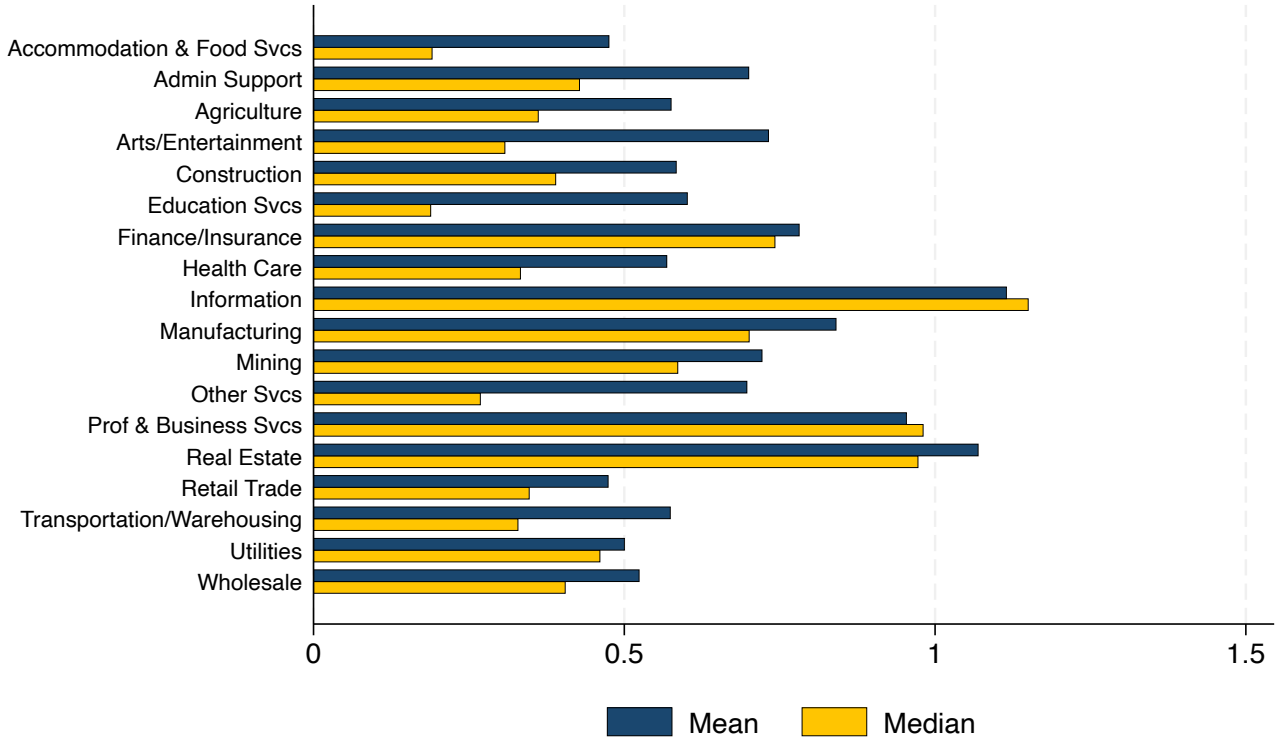
This figure compares aggregate employment in Revelio and Compustat for the U.S. public-firm sample from 2010 to 2024. The solid blue line reports the total global Revelio jobs each year, and the dashed red line reports aggregate Compustat employment for the same matched U.S. public firms.

Figure A.2. Industry-Level Employment Coverage Ratios (Employment-weighted)



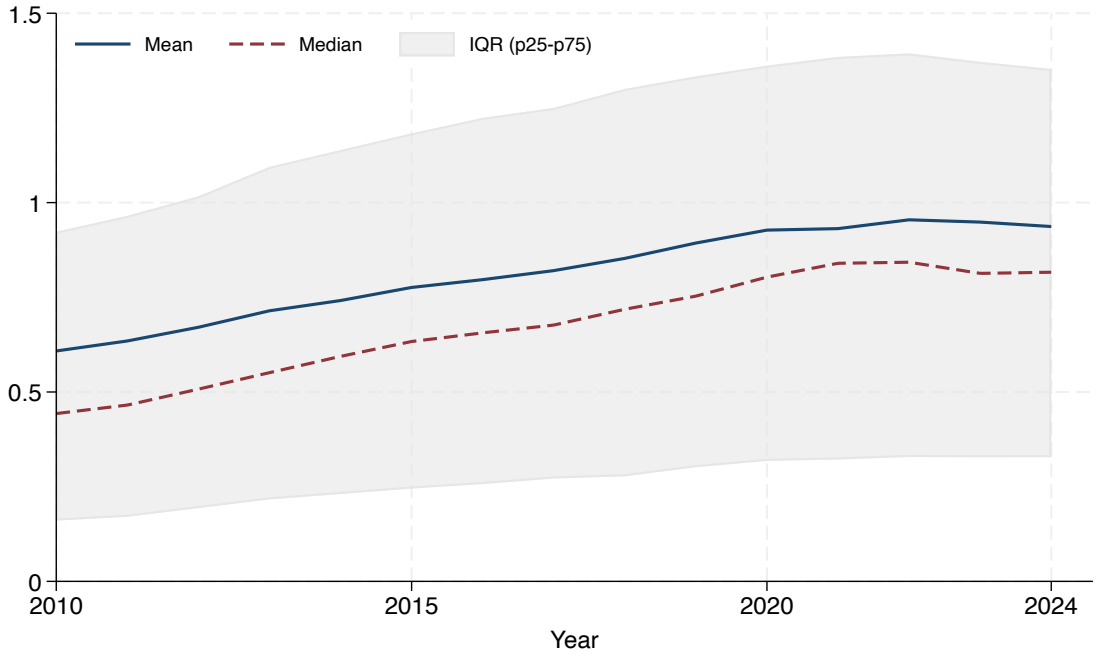
This figure reports industry-level employment coverage ratios for all matched U.S. public firms from 2010 to 2024. For each sector-year (based on 2-digit NAICS industry codes), the coverage ratio is defined as the sum of all global Revelio jobs divided by the sum of Compustat employment. Bars report the mean and the median of the annual industry-level coverage ratio over the sample period.

Figure A.3. Industry-Level Employment Coverage Ratios (Equal-weighted)



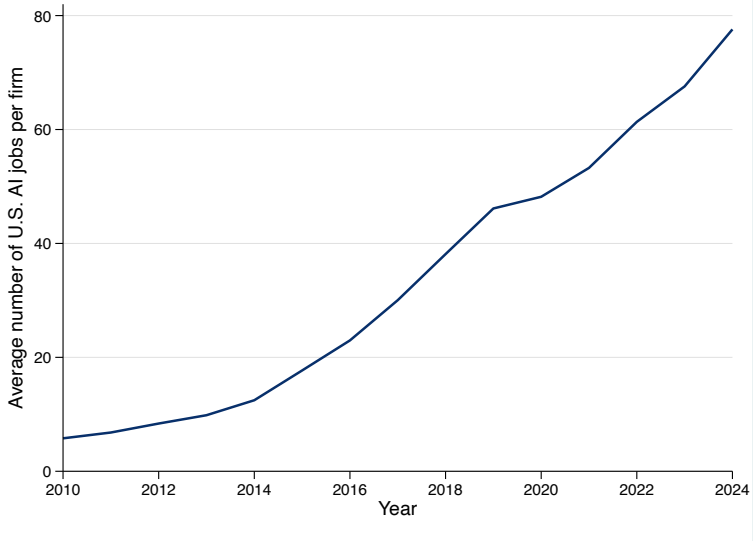
This figure reports equal-weighted industry-level statistics for the firm-level employment coverage ratio for all matched U.S. public firm-years from 2010 to 2024. The coverage ratio is defined as total global Revelio jobs divided by Compustat employment and is computed at the firm-year level. The ratio is winsorized within year at the 1st and 99th percentiles before computing industry statistics. Bars report the mean and median coverage ratio across all firm-year observations within each sector (based on 2-digit NAICS industry codes) over the sample period.

Figure A.4. Firm-Level Employment Coverage Ratio by Year

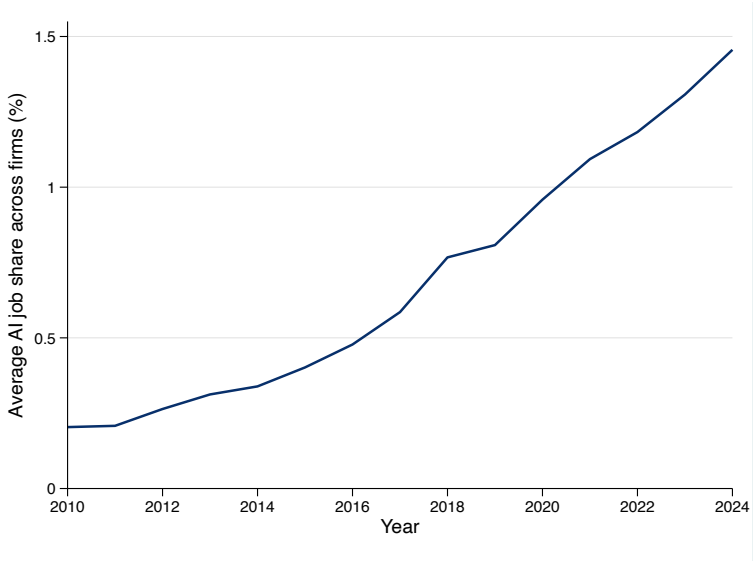


This figure reports annual distributional statistics for the firm-level employment coverage ratio for all matched U.S. public firm-years from 2010 to 2024. The coverage ratio is defined as total global Revelio jobs divided by Compustat employment and is computed at the firm-year level. The ratio is winsorized within year at the 1st and 99th percentiles before computing annual statistics. The solid blue line reports the cross-sectional mean by year, the dashed red line reports the cross-sectional median by year, and the gray shaded area denotes the interquartile range, defined by the 25th and 75th percentiles.

Figure A.5. Average Firm-Level AI-Related Jobs and AI Job Shares



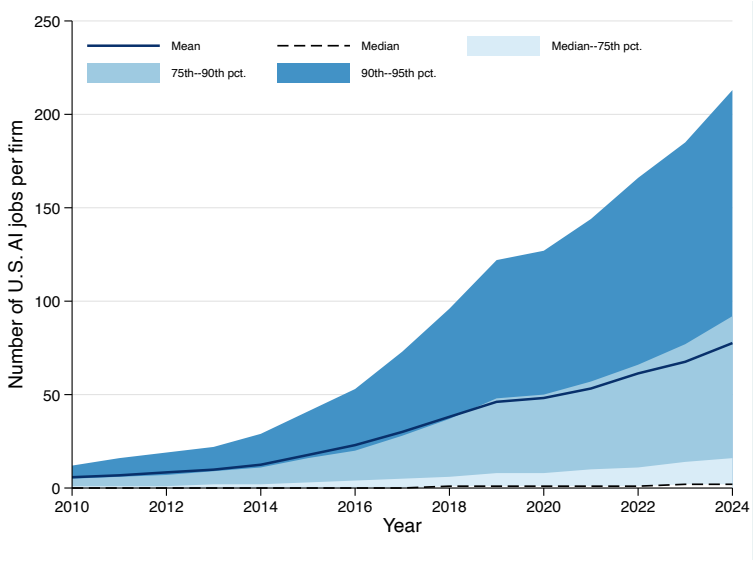
(a) AI jobs



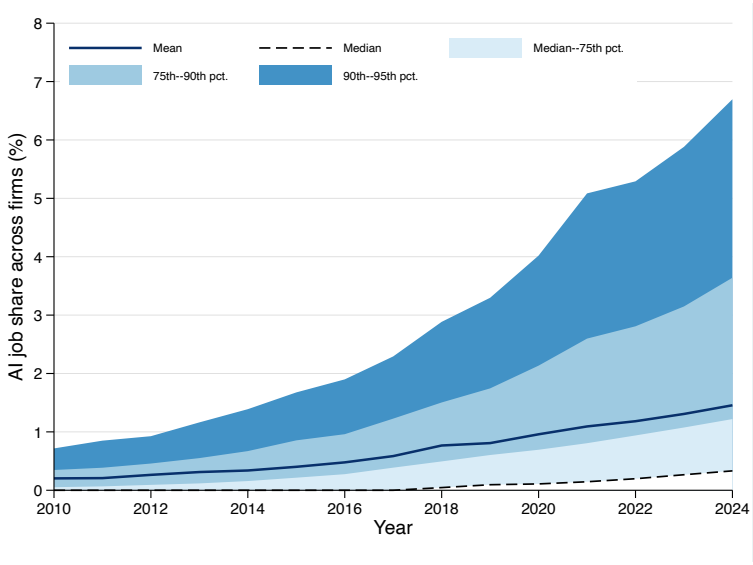
(b) AI shares

This figure reports average firm-level AI-related employment from 2010 to 2024 for the U.S. public-firm sample. Panel (a) plots the average number of AI-related U.S. jobs per firm. Panel (b) plots the average firm-level AI job share, defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. Averages are computed across firms within each year. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.6. Distribution of Firm-Level AI-Related Jobs and AI Job Shares



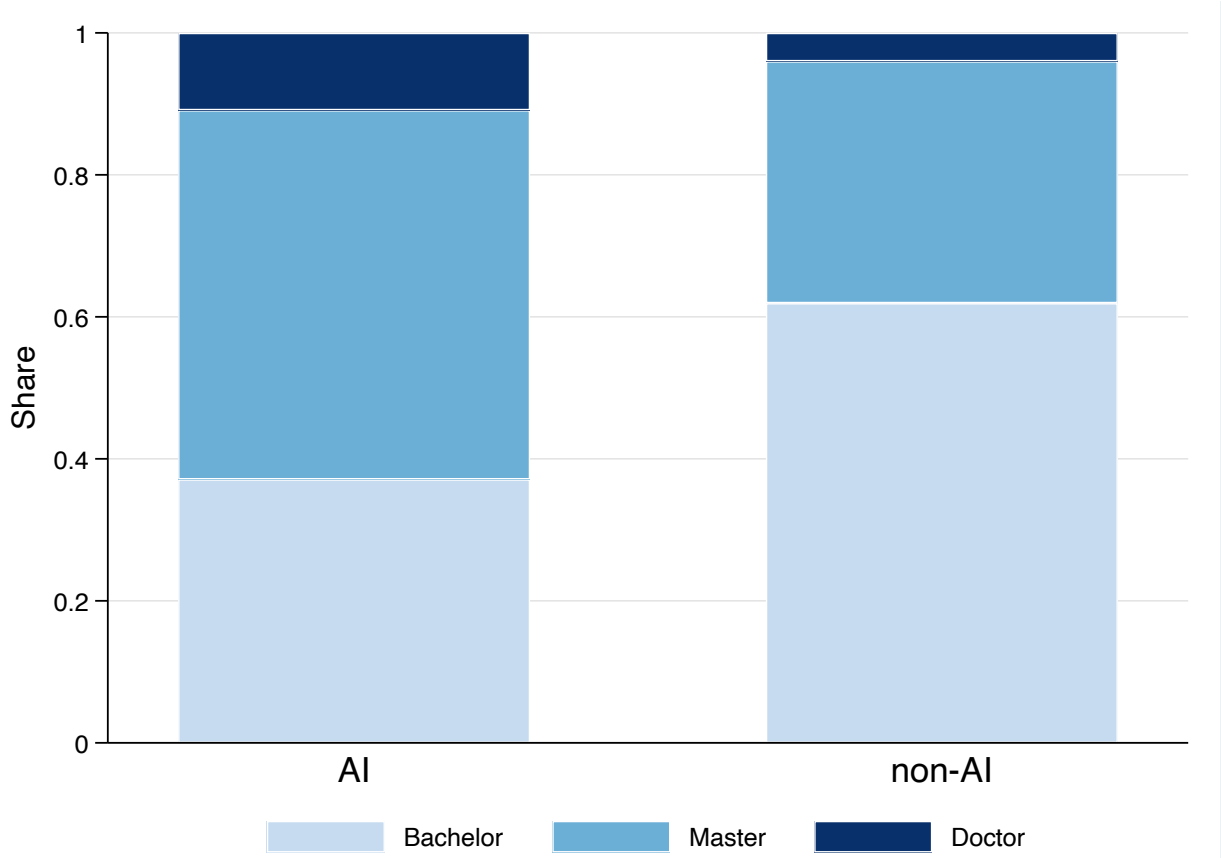
(a) AI jobs



(b) AI shares

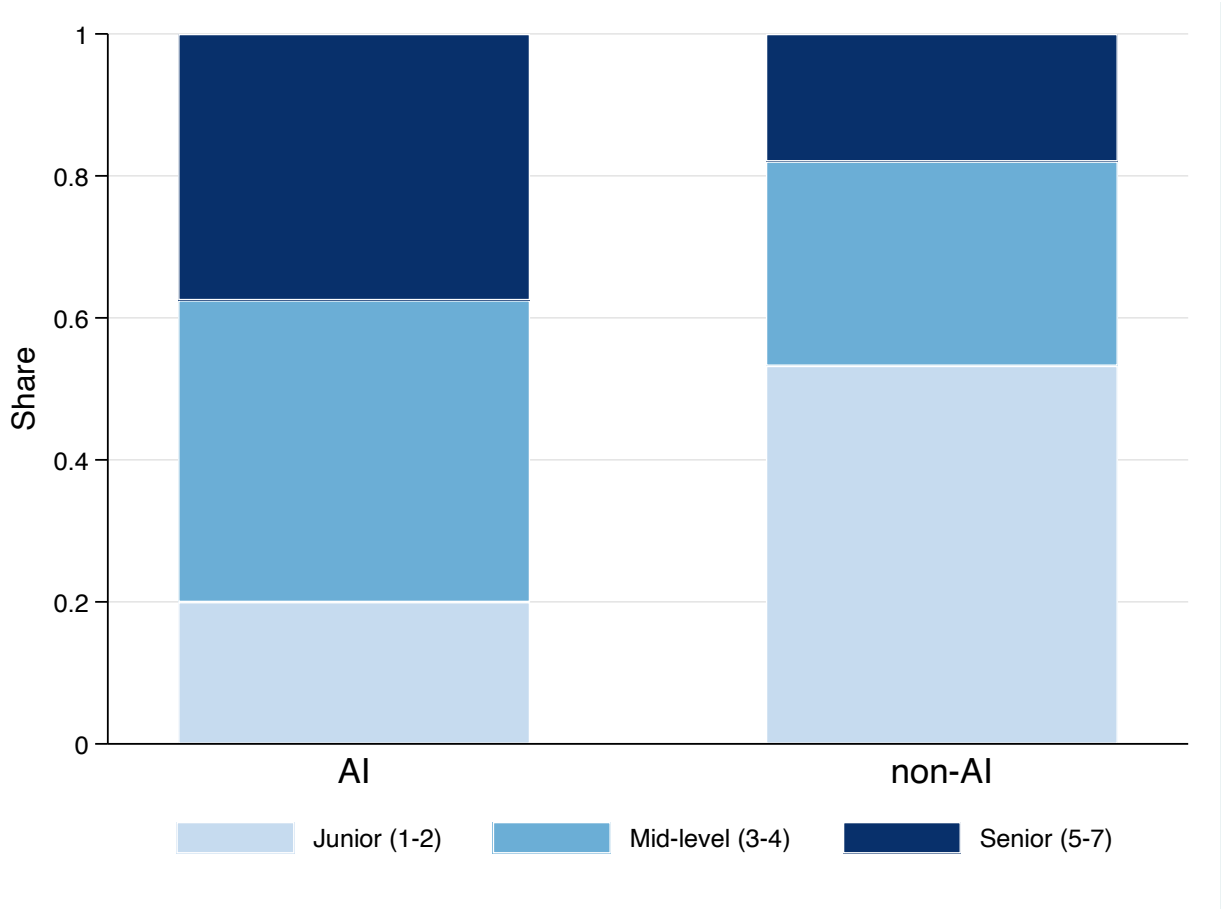
This figure reports the distribution of firm-level AI-related employment from 2010 to 2024 for the U.S. public-firm sample. Panel (a) plots the number of AI-related U.S. jobs per firm. Panel (b) plots the firm-level AI job share, defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. Statistics are computed across firms within each year. The shaded areas report the median–75th percentile, 75th–90th percentile, and 90th–95th percentile ranges. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.7. Educational Attainment of AI-Related versus Non-AI-Related Workers



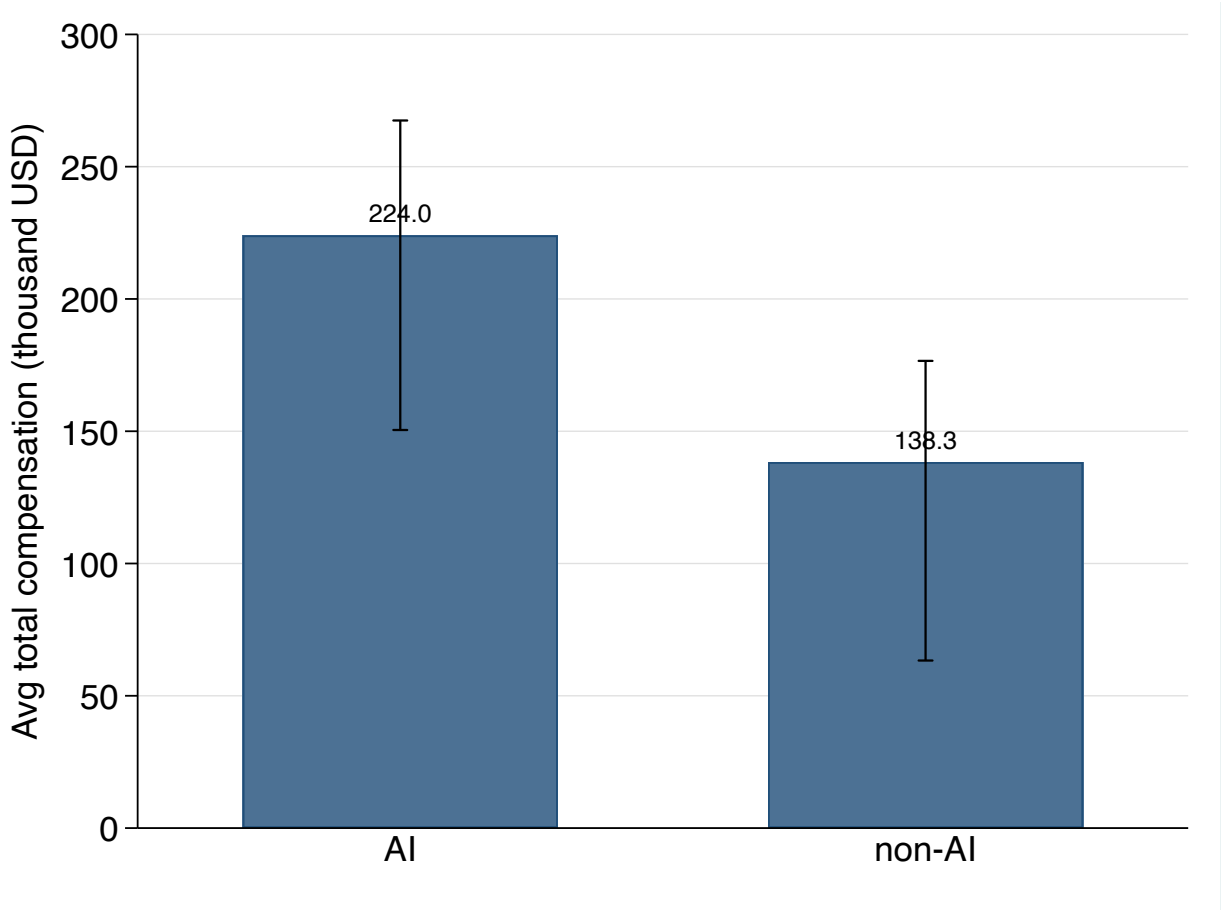
This figure plots the highest educational attainment of AI-related and non-AI-related workers among U.S. jobs at the job level. Each bar reports the share of U.S. positions within AI-related or non-AI-related jobs that falls into the corresponding degree category in 2024 for the U.S. public-firm sample. Bachelor includes bachelor's and associate degrees, Master includes master's degrees and MBAs, and Doctor includes doctoral degrees. Positions with missing or unrecognized education and high-school degrees are excluded. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.8. Seniority Distribution of AI-Related versus Non-AI-Related Workers



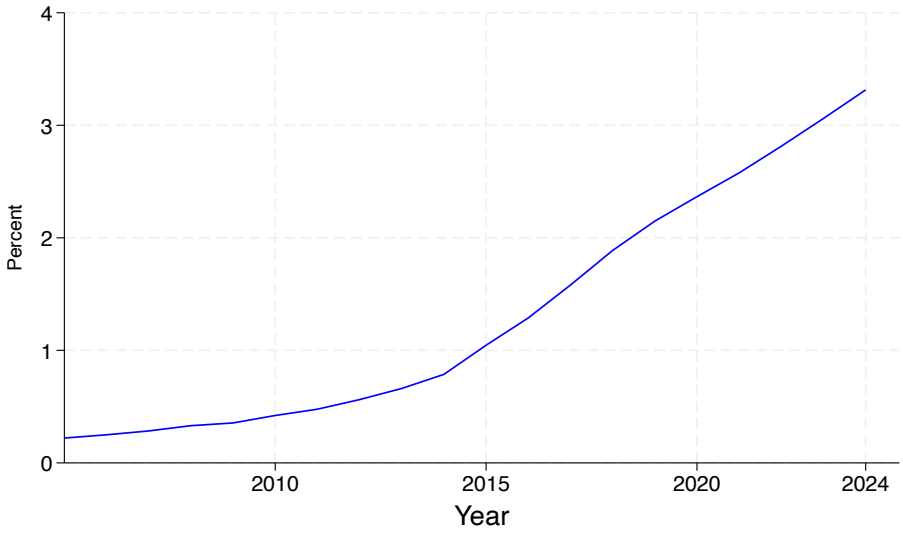
This figure plots the seniority distribution of AI-related and non-AI-related workers among U.S. jobs at the job level. Each bar reports the share of U.S. positions within AI-related or non-AI-related jobs that falls into the corresponding seniority category in 2024 for the U.S. public-firm sample. Seniority is based on Revelio's seven-level seniority measure: 1 is Entry Level, 2 is Junior Level, 3 is Associate Level, 4 is Manager Level, 5 is Director Level, 6 is Executive Level, and 7 is Senior Executive Level. We group levels 1–2 as Junior, levels 3–4 as Mid-level, and levels 5–7 as Senior. Positions with missing or unrecognized seniority are excluded. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.9. Average Total Compensation of AI-Related versus Non-AI-Related U.S. Jobs



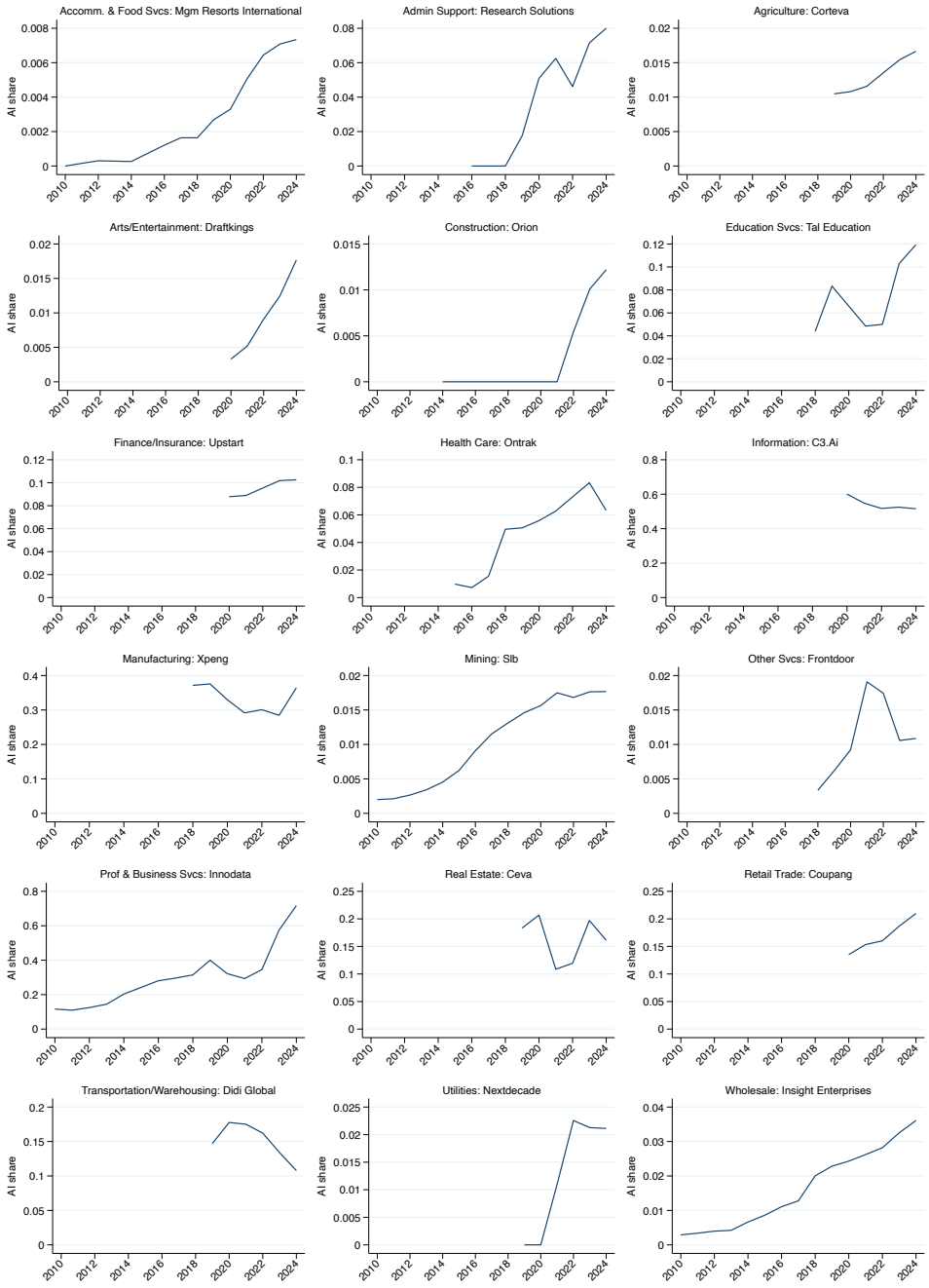
This figure plots average total compensation for AI-related and non-AI-related U.S. jobs in 2024 for the U.S. public-firm sample. Total compensation for each position is imputed by Revelio. Bars report aggregate total compensation divided by the number of jobs, computed separately for AI-related and non-AI-related U.S. jobs, and expressed in thousands of U.S. dollars. Vertical lines show the 25th and 75th percentiles of job-level total compensation within AI-related and non-AI-related U.S. jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.10. Wage-Weighted Share of AI Jobs



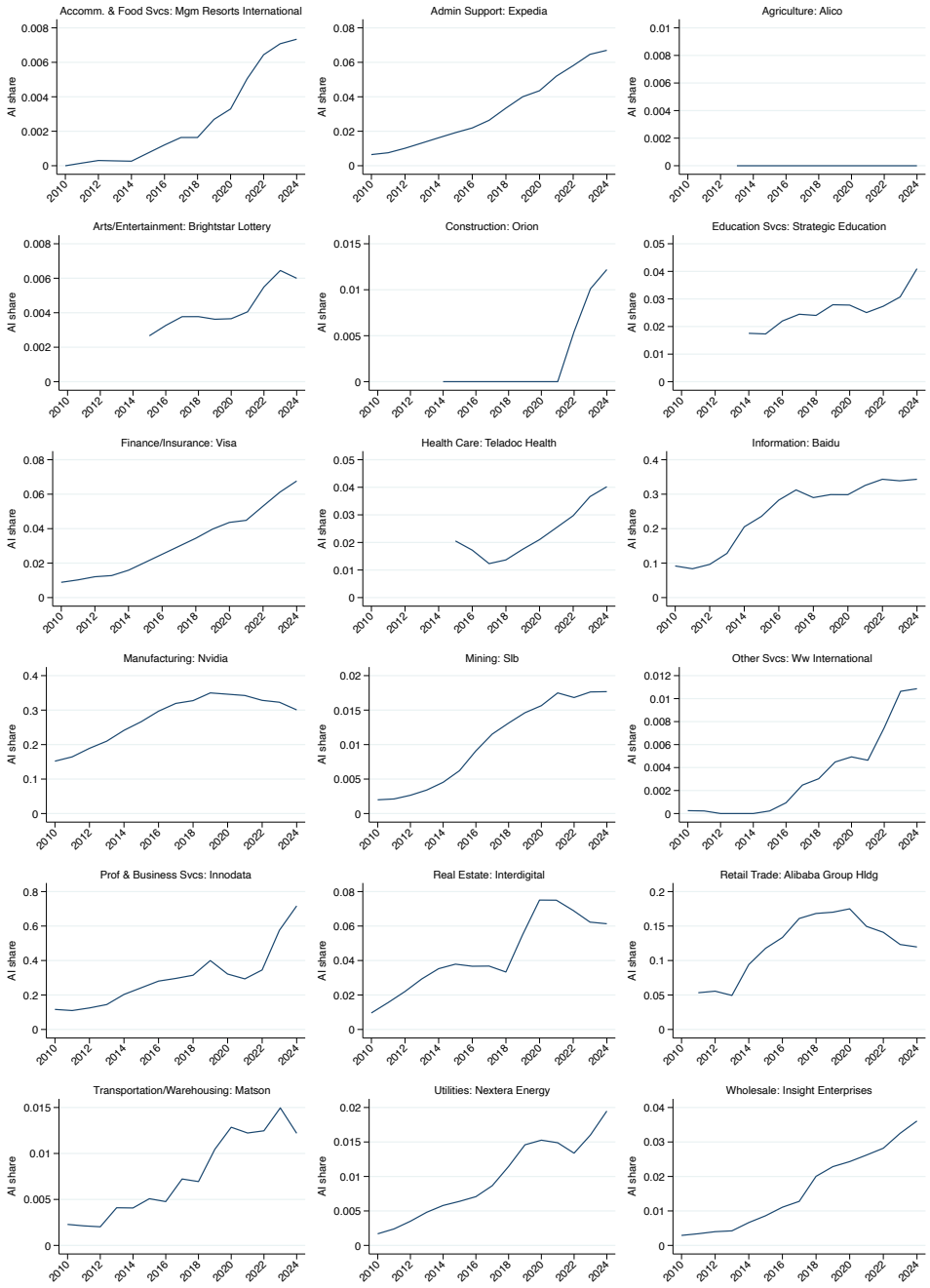
This figure plots the wage-weighted share of AI-related jobs among U.S. jobs in the U.S. public-firm sample from 2010 to 2024. We weight jobs by their annual compensation imputed by Revelio. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.11. AI Share of 2024 Industry Leaders with Five-Year Continuous Histories



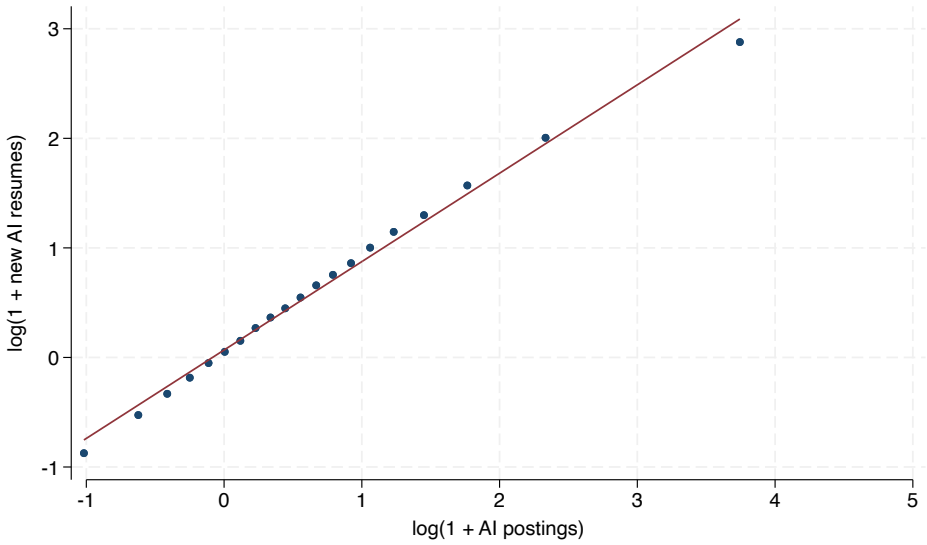
This figure plots the time series of AI share for the top firm in each sector (based on 2-digit NAICS industry codes) from 2010 to 2024 for the U.S. public-firm sample. The top firm in each industry is defined as the public firm with the highest AI share in 2024 among firms observed continuously over the five-year period ending in 2024 and with at least 50 U.S. workers in Revelio and 50 employees in Compustat. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.12. AI Share of 2024 Industry Leaders with Ten-Year Continuous Histories



This figure plots the time series of AI share for the top firm in each sector (based on 2-digit NAICS industry codes) from 2010 to 2024 for the U.S. public-firm sample. The top firm in each industry is defined as the public firm with the highest AI share in 2024 among firms observed continuously over the ten-year period ending in 2024 and with at least 50 U.S. workers in Revelio and 50 employees in Compustat. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure A.13. AI-Related Job Postings versus New AI Resumes



This figure shows a residualized binscatter of log(1 + new AI resume jobs) against log(1 + AI-related job postings) at the firm-year level, 2010–2024. A new AI resume job is a Revelio position that begins at firm f in year t and whose text matches an AI skill term; an AI-related posting is a Lightcast job ad posted by firm f in year t that lists one of the same AI skills. Both sides are identified by a common skill list: a skill enters the list in the first year it appears in the postings data and is applied to both sources from that year onward, and within each firm-year only skills observed in both the firm's postings and its new resumes are used, so the two measures reflect an identical, contemporaneous definition of AI. Both series are residualized on year fixed effects, the log of the firm's total job postings, and the log of the firm's total U.S. resume jobs. The sample consists of firm-years in the U.S. public-firm sample observed in both Lightcast and Revelio data. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.1. Employment Coverage Rates by Sample Characteristics

Sample	N	Mean	Std. Dev.	p25	p50	p75
U.S. headquartered	58,260	86.0	72.3	30.0	71.9	128.8
Non-U.S. headquartered	14,329	63.1	66.6	10.2	43.2	100.5
U.S.-majority workers	51,411	86.5	70.7	31.1	73.4	129.4
Non-U.S.-majority workers	21,178	69.3	72.9	13.3	49.3	107.4

This table reports firm-year employment coverage rates for U.S. public firms from 2010 to 2024. The coverage rate is defined as total global Revelio jobs divided by Compustat employment for the same firm-year and is expressed in percentages (this coverage rate is not the same as the coverage ratio filter we used). The rate is winsorized within year at the 1st and 99th percentiles before computing statistics. All reported moments are computed across firm-years. Rows report separate sample splits and are not cumulative. U.S.-headquartered firms are firms classified as headquartered in the United States. U.S.-majority worker firms are firm-years for which more than half of workers observed in Revelio are located in the United States.

Table A.2. Summary Statistics for Compustat Employment and Revelio Jobs

	Compustat Employment			Revelio Jobs		
Statistic	2010	2018	2024	2010	2018	2024
p1	5	4	5	1	1	2
p5	19	13	16	4	5	8
p25	193	149	151	60	81	98
p50	1,164	1,173	1,166	353	554	644
p75	6,200	7,400	7,500	1,780	3,186	3,803
p95	58,300	69,000	68,852	16,647	28,652	34,207
Mean	12,593	14,355	14,517	3,903	6,519	7,648
St.d	49,620	55,945	58,819	15,808	24,688	30,869
Skewness	19	17	17	13	10	13
Kurtosis	682	528	476	259	156	277
N	4,783	4,716	4,562	4,783	4,716	4,562

This table reports summary statistics for Compustat employment and total global Revelio jobs for the matched U.S. public firms in 2010, 2018, and 2024. Revelio jobs are the annual count of all jobs observed in Revelio for each firm-year. Statistics are computed across firms within each year and are unweighted.

Table A.3. Correlation Between Revelio and Compustat Employment by Employment Coverage Ratio

Coverage bin	N	Number of Firms	Total	Firm FE
All	72,589	9,086	0.610	0.618
< 0.05	6,894	1,278	0.674	0.328
0.05–0.2	9,137	1,898	0.972	0.469
0.2–1	33,602	5,321	0.904	0.873
1–2	20,224	4,011	0.987	0.937
2–3	1,672	795	0.998	0.991
≥ 3	1,060	361	0.609	0.376

This table reports Pearson correlations between total global Revelio jobs and Compustat employment for all matched U.S. public firm-years from 2010 to 2024. The employment coverage ratio is defined as total global Revelio workers divided by Compustat employment for the same firm-year. Coverage bins are assigned at the firm-year level. "N" reports the number of firm-year observations, and "Number of Firms" reports the number of unique firms in each bin. "Total" stands for the cross-sectional correlation that is computed in employment levels. "Firm FE" stands for the correlation that is computed after demeaning both employment measures by firm within each coverage-bin sample.

Table A.4. Correlation Between Revelio and Compustat Employment by Firm Size

Sample	N	Number of Firms	Total	Firm FE
All	72,589	9,086	0.610	0.618
Revelio Worker ≥ 50	57,404	6,983	0.607	0.623
Revelio Worker < 50	15,185	3,157	-0.016	0.056
Compustat employment ≥ 50	63,279	7,842	0.607	0.618
Compustat employment < 50	9,310	2,269	0.037	0.505

This table reports Pearson correlations between total global Revelio jobs and Compustat employment for all matched U.S. public firm-years from 2010 to 2024. Rows split the sample by firm size, using either total Revelio workers or Compustat employment for the same firm-year. "N" reports the number of firm-year observations, and "Number of Firms" reports the number of unique firms in each row. The cross-sectional correlation is computed in levels. The firm FE correlation is computed after demeaning total global Revelio jobs and Compustat employment by firm within each row's sample.

Table A.5. Firm Counts by Sample Filter

Year	Matched	Coverage $(0.05, 3)$	U.S. headquartered	Revelio U.S. workers ≥ 50	U.S.-majority	Compustat employment ≥ 50	All filters
2010	4,783	4,099	4,040	3,309	3,563	4,264	2,482
2011	4,747	4,102	3,972	3,308	3,514	4,231	2,471
2012	4,761	4,160	3,947	3,290	3,489	4,167	2,424
2013	4,928	4,350	4,078	3,391	3,617	4,244	2,491
2014	4,971	4,415	4,095	3,453	3,649	4,255	2,526
2015	4,866	4,331	3,985	3,430	3,548	4,184	2,483
2016	4,790	4,282	3,875	3,376	3,463	4,138	2,445
2017	4,727	4,235	3,807	3,349	3,387	4,102	2,386
2018	4,716	4,228	3,774	3,373	3,339	4,099	2,391
2019	4,650	4,168	3,685	3,371	3,267	4,078	2,362
2020	4,953	4,458	3,857	3,517	3,401	4,283	2,434
2021	5,196	4,687	4,032	3,690	3,535	4,541	2,550
2022	5,083	4,597	3,911	3,654	3,424	4,450	2,479
2023	4,856	4,382	3,715	3,473	3,218	4,252	2,335
2024	4,562	4,141	3,487	3,295	2,997	3,991	2,179

This table reports the number of matched U.S. public firms by year under alternative sample filters. "Matched" firms are firms in the matched Revelio–Compustat sample. The "Coverage" filter keeps firm-years with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year. "U.S.-headquartered" firms are firms classified as headquartered in the United States. "Revelio U.S. workers ≥ 50" keeps firm-years with at least 50 U.S.-based workers observed in Revelio. "U.S.-majority" firms are firm-years for which more than half of workers observed in Revelio are located in the United States. "Compustat employment ≥ 50" keeps firm-years with at least 50 employees in Compustat. Each filter column applies the listed condition separately. The final column applies all listed filters jointly.

Table A.6. Skills with Highest AI-Relatedness Measures in Lightcast Job Postings

#	Skills	AI-relatedness Score
1	Machine Learning	1.000
2	Artificial Intelligence	1.000
3	Natural Language Processing (NLP)	1.000
4	Computer Vision	1.000
5	Zero Shot Learning	1.000
6	Rweka	1.000
7	AgentGPT	1.000
8	Doc2vec	0.995
9	Few Shot Learning	0.991
10	fastText	0.989
11	Hugging Face Transformers	0.989
12	Confusion Matrix	0.986
13	Hugging Face (NLP Framework)	0.985
14	CrewAI	0.985
15	Seq2Seq	0.983
16	AI-Generated Code	0.983
17	Hyperparameter Optimization	0.981
18	Agentic Systems	0.980
19	Standard ML Library	0.977
20	PyTorch Lightning	0.977
21	Amazon Rekognition	0.977
22	Neural Architecture Compression	0.976
23	AWS Inferentia	0.976
24	Microsoft Cognitive Toolkit (CNTK)	0.976
25	Dbscan	0.976
26	Word2Vec Models	0.975
27	CatBoost (Machine Learning Library)	0.972
28	Caffe2	0.972
29	Apache MXNet	0.971
30	Feature Learning	0.969
31	Reinforcement Learning from Human Feedback (RLHF)	0.969
32	OpenAI Gym	0.966
33	BERT (NLP Model)	0.966
34	PyTorch (Machine Learning Library)	0.966
35	Adversarial Machine Learning	0.964
36	Vowpal Wabbit	0.963
37	Attention Mechanisms	0.962
38	Retrieval Augmented Generation	0.962
39	Open Neural Network Exchange (ONNX)	0.961
40	Long Short-Term Memory (LSTM)	0.961
41	Weaviate	0.960
42	Small Language Model	0.960
43	Text Summarization	0.957
44	Part-of-Speech Tagging	0.957
45	DeepSpeech	0.957
46	Autoencoders	0.956
47	SpaCy (NLP Software)	0.956
48	Word Embedding	0.956
49	Neural Architecture Search (NAS)	0.956
50	Keras (Neural Network Library)	0.955

This table lists the top skills in the Lightcast data ranked by the skill-level AI measure w_s^{AI}. For each skill, we report the fraction of jobs requiring that skill that also require one of the four core AI skills—artificial intelligence, machine learning, computer vision, and natural language processing. Only skills that appear in at least 50 job postings are included.

Table A.7. False-Positive Rates for AI-Related Jobs by Firm Size

Tercile	N	FPs	FP rate
Small	5,084	278	5.47%
Mid	5,084	172	3.38%
Large	5,084	103	2.03%

This table reports false-positive rates for AI-related U.S. jobs by firm-size tercile. The sample covers 2010–2024, excludes internships and part-time positions, and is restricted to matched U.S. public firm-years with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year. Firm-size terciles are assigned within year using total global Revelio job counts. All AI-related jobs at small-tercile firms are included. The mid- and large-firm samples are random draws from their respective tercile job-year pools, with draws matched to the small-firm sample by year. "N" reports the number of sampled position-years, "FPs" reports the number classified as false positives, and "FP rate" shows the percentage of false-positive jobs.

Table A.8. False-Positive Rates for AI-Related Jobs by AI Subcategory

Subcategory	N	FPs	FP rate
AI	5,000	88	1.76%
Agentic AI	5,000	35	0.70%
CV	5,000	62	1.24%
Gen AI	5,000	24	0.48%
ML	5,000	80	1.60%
NLP	5,000	134	2.68%
Other	5,000	18	0.36%

This table reports false-positive rates from random samples of AI-related U.S. jobs by AI subcategory. For each subcategory, the sample includes 5,000 jobs from 2010 to 2024 in the U.S. public-firm sample, excluding internships and part-time positions. Subcategories are assigned using keyword matches in job titles or job descriptions. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year. "N" reports the number of sampled jobs, "FPs" reports the number classified as false positives, and "FP rate" shows the percentage of false-positive jobs.

Table A.9. Annual AI-Related Job Shares by Technology Category

Year	(1) Share AI All (%)	(2) Share General AI (%)	(3) Share ML (%)	(4) Share NLP (%)	(5) Share CV (%)	(6) Share Gen AI (%)	(7) Share Agentic AI (%)	(8) Share AI Other (%)
2010	0.2508	0.1167	0.1347	0.0338	0.0232	0.0083	0.0015	0.0140
2011	0.2753	0.1292	0.1507	0.0360	0.0249	0.0094	0.0017	0.0155
2012	0.3221	0.1517	0.1799	0.0418	0.0295	0.0110	0.0018	0.0191
2013	0.3733	0.1771	0.2141	0.0473	0.0320	0.0125	0.0021	0.0228
2014	0.4491	0.2152	0.2643	0.0563	0.0361	0.0156	0.0025	0.0284
2015	0.5968	0.2908	0.3594	0.0722	0.0460	0.0225	0.0035	0.0411
2016	0.7390	0.3769	0.4470	0.0851	0.0508	0.0294	0.0045	0.0539
2017	0.9195	0.4835	0.5591	0.1039	0.0594	0.0395	0.0061	0.0694
2018	1.1027	0.6010	0.6646	0.1232	0.0670	0.0544	0.0081	0.0893
2019	1.2697	0.7162	0.7527	0.1385	0.0724	0.0771	0.0115	0.1115
2020	1.4037	0.8252	0.8079	0.1464	0.0762	0.1075	0.0162	0.1321
2021	1.5426	0.9513	0.8431	0.1462	0.0758	0.1591	0.0244	0.1558
2022	1.7029	1.1125	0.8556	0.1403	0.0735	0.2450	0.0385	0.1758
2023	1.8845	1.2958	0.8610	0.1390	0.0694	0.3527	0.0596	0.1936
2024	2.0990	1.5008	0.8710	0.1367	0.0640	0.4514	0.0921	0.2106

This table reports annual AI-related job shares from 2010 to 2024 for the U.S. public-firm sample. All shares are computed using U.S. jobs only. Share is defined as the fraction of U.S. jobs whose title or description contains at least one AI keyword. "Share AI all (%)" uses the full AI keyword dictionary. The remaining columns report shares for six AI technology categories: general AI, machine learning (ML), natural language processing (NLP), computer vision (CV), generative AI (Gen AI), and agentic AI. The "Share AI other" category includes AI-related jobs that are not assigned to any of the six named technology subcategories. A job is counted at most once within a category even if it matches multiple keywords in that category, but categories are not mutually exclusive across columns. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.10. Top 50 Keywords with the Most AI Jobs in 2024

Rank	Keyword	Job Counts	Rank	Keyword	Job Counts
1	AI	219,613	26	feature engineer	3,176
2	machine learning	66,932	27	ai agents	3,121
3	ML	59,366	28	prompt engineering	2,974
4	generative ai	23,930	29	xgboost	2,765
5	ai/ml	23,672	30	matplotlib	2,729
6	LLM	18,986	31	chatgpt	2,650
7	artificial intelligence	18,549	32	conversational ai	2,618
8	genai	17,995	33	mlflow	2,431
9	NLP	10,046	34	retrieval-augmented generation	2,358
10	deep learning	9,959	35	random forest	2,337
11	LLMs	9,664	36	vertex ai	2,195
12	computer vision	7,313	37	sentiment analysis	2,084
13	language model	6,717	38	aws sagemaker	1,982
14	large language model	5,508	39	reinforcement learning	1,910
15	numpy	5,483	40	microsoft copilot	1,897
16	tensorflow	4,982	41	BERT	1,871
17	mlops	4,891	42	machine learning technique	1,842
18	agentic ai	4,715	43	ai agent	1,826
19	pytorch	4,477	44	github copilot	1,786
20	natural language processing	4,088	45	azure openai	1,783
21	scikit-learn	4,020	46	ml/ai	1,741
22	langchain	3,823	47	ai innovation	1,646
23	machine learning algorithm	3,488	48	gpt-4	1,628
24	neural network	3,331	49	scipy	1,544
25	ai research	3,313	50	hugging face	1,458

This table reports the 50 AI keywords with the largest number of job-level matches among U.S. jobs in 2024 for the U.S. public-firm sample. A job is counted once for a given keyword if the keyword appears in either the job title or the job description. A job can match multiple keywords, so counts are not mutually exclusive. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.11. Top 10 Keywords with the Most AI Jobs by AI Subcategory in 2024

AI		**ML**	
Keyword	Job Counts	Keyword	Job Counts
AI	219,613	machine learning	66,932
artificial intelligence	18,549	ML	59,366
ai research	3,313	deep learning	9,959
ai innovation	1,646	numpy	5,483
ai security	951	tensorflow	4,982
ai development	933	mlops	4,891
ethical ai	573	pytorch	4,477
artificial general intelligence	451	scikit-learn	4,020
explainable ai	403	machine learning algorithm	3,488
ai safety	356	neural network	3,331
NLP		**CV**	
Keyword	Job Counts	Keyword	Job Counts
NLP	10,046	computer vision	7,313
natural language processing	4,088	object detection	1,115
conversational ai	2,618	image recognition	625
sentiment analysis	2,084	opencv	624
BERT	1,871	CNNs	386
AIML	1,257	image segmentation	207
spacy	772	3d reconstruction	189
nltk	752	scene understanding	143
NLTK	703	pose estimation	111
natural language understanding	668	object recognition	101
Gen AI		**Agentic AI**	
Keyword	Job Counts	Keyword	Job Counts
generative ai	23,930	agentic ai	4,715
LLM	18,986	langchain	3,823
genai	17,995	ai agents	3,121
LLMs	9,664	ai agent	1,826
language model	6,717	langgraph	1,053
large language model	5,508	copilot studio	941
prompt engineering	2,974	agentic system	327
chatgpt	2,650	autogen	276
retrieval-augmented generation	2,358	intelligent agent	255
microsoft copilot	1,897	crewai	249

This table reports the 10 AI keywords with the largest number of job-level matches in 2024 within each of six AI subcategories. The sample includes U.S. jobs in the U.S. public-firm sample. A job is counted once for a given keyword if the keyword appears in either the job title or the job description. A job can match multiple keywords, so counts are not mutually exclusive. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.12. Examples of AI-Related Jobs by AI Subcategory

Example	Job title	Job description
Panel A: General AI		
1	Director, Medical Affairs Oncology **AI Innovation**	Driving **AI Innovation**: Leading the integration of **artificial intelligence** into Medical Affairs to optimize workflows, enhance decision-making, and deliver impactful insights. Empowering Oncology Teams: Providing **AI**-driven tools and solutions that enable the oncology team to address complex challenges and improve patient outcomes. Strategic Leadership: Developing a comprehensive **AI** roadmap that aligns with business goals and ensures scalable and sustainable solutions. Evidence Generation: Leveraging **AI** for real-world evidence (RWE) analysis and predictive analytics to support clinical and therapeutic advancements. Collaboration Across Functions: Partnering with cross-functional teams, including R&D, regulatory, and commercial, to align **AI** initiatives with broader organizational objectives. Stakeholder Engagement: Building relationships with internal and external stakeholders to drive the adoption and trust of **AI**-driven solutions in oncology. **Ethical AI** Implementation: Ensuring all **AI** solutions are compliant, unbiased, and focused on improving patient care.
2	Information Architect, UI/UX Design Lead	Transforming user complexity into intuitive digital experiences that deliver measurable results. Selected to transition from 3M to Solventum, helping build the new company's digital product foundation. • Leading Vision-to-Execution Design: Owning design strategy from concept to launch, aligning the team on visually compelling interfaces for the best user experience. • Bridging Design & Development: Facilitating collaboration across teams to ensure the design vision is translated flawlessly into shippable products. • Designing Intuitive User Journeys: Architecting the underlying structure of digital products, championing user needs to boost engagement and retention. • Integrating **AI** in the Design Workflow: Leveraging **AI** to automate tasks, empowering the team to accelerate prototyping and focus on innovation.
3	Senior Manager, Field & Account Based Marketing	Oversees Zillow's largest execution-focused B2B marketing team, leading Field and Account-Based Marketing strategy and delivery for top partner segments. Owns strategy and leadership for Field and Account-Based Marketing programs that drive product adoption across Zillow's mortgage offerings and real estate CRM platform, contributing to record-setting month-over-month transaction growth. Hosted Zillow's 2024 National Home Buyer Webinar Series, leveraging real estate expertise and on-camera experience to strengthen Zillow's consumer education presence. Embeds **AI** strategy into team operations (e.g., **AI** voiceovers, market updates, and enablement content) to accelerate execution and improve team efficiency. Restructured team model to scale with business needs, improved cross-functional alignment, and built internal leadership capacity. Is recognized across Marketing and Sales leadership as a performance-focused, high-impact people leader.
4	Senior Director, Finance Data Strategy Leader	Lead the design and implementation of the global Financial Data Strategy to modernize finance through data integration, automation, and advanced analytics. Build and scale a financial data platform using cutting-edge technologies, enabling seamless access to ERP and non-ERP data for real-time reporting, deep analytical capabiities, and decision support. Drive Digital Transformation initiatives focused on process optimization, **AI** integration, and automation of financial data pipelines. Establish Data Governance frameworks to ensure data quality, security, and compliance across the finance organization. Develop and lead a Finance Upskilling program to enhance capabilities in data literacy, analytics, and technology adoption. Define KPIs, success metrics, and governance models to measure impact, improve operational efficiency, and support a culture of continuous innovation.
5	VP, Retail and Digital Technology	Leading Product and Engineering functions across our Digital and Shop technologies as a proven technology and product leader who led a 30+ person team to deliver digital solutions that fueled 33% revenue growth, scaling the business to $1.28B in just 15 months. Drove mobile platform adoption with 75% loyalty usage and 15% channel growth, while aligning closely with C-level stakeholders to unlock $120M in new revenue through mobile ordering. Modernized product delivery by introducing agile frameworks, dual-track workflows, and CI/CD pipelines-accelerating time to market and improving delivery confidence. Applied **AI** and automation to cut support call volume by 20% and reduce handle time by 30%. Built high-performing cross-functional teams, overhauled legacy systems to reduce operational costs by 30%, and implemented a hybrid talent model that lowered development costs by 20%. Known for delivering scalable strategy, strong execution, and measurable business value.
6	Senior Research Scientist	Research on Digital Humans, 3D Generative Models, and Embodied **AI**.
7	Vice President, Product Manager	Wealth Plan Driving innovation in digital wealth management by building and launching Wealth Plan, a personalized financial planning experience used by millions of clients. Partnered with engineering, design, and data science to deliver seamless goal-setting, **AI**-driven insights, and intuitive user experiences that empower clients to plan, track, and achieve their long-term financial goals. Skilled at transforming complex financial products into simple, client-first digital solutions at scale.
8	SVP, Chief Data Officer & Enterprise Innovation	Driving the business process optimization governance frameworks that guide and monitor the relationship between Xerox and Business Process Outsourcing provider(s), and designing and developing a holistic data and digital strategy across the organization. • Leading **AI** & Data Governance Center of excellence • Developing data and digital strategy, including master data management and governance processes to execute across operational units. • Creating a roadmap for an advanced data infrastructure, including a data warehouse, data lake, ingestion pipeline, and accessible APIs, to drive efficiencies across the organization. • Driving Business Process Outsourcing (BPO) strategy and governance,
9	Innovation Manager	Develop virtual reality + **artificial intelligence** solutions to improve the interactive experience of intelligent VR scenes.
10	Lead Product Designer	• Lead Product Designer: • Contextual **AI** recommendations • Compute Usage Optimizer • Customer Support Homepage widget • **AI**-powered Logs Analysis sidekick • Materialized Views
Panel B: Machine Learning		

Continued on next page

Table A.12 continued

Example	Job title	Job description
1	Measurement and Insights	Analyzed qualitative and quantitative data in Databricks and Tableau to identify the impact of sponsorship at Live Nation festivals and events. Distilled and organized information for decks from raw data for Fortune 500 partners and event deliverables. Built a **random forest** feature importance model, using Python to identify factors highly correlated with high fan satisfaction, which informed future resource allocation and brand marketing strategy
2	NLP Data Scientist & **ML** Engineer	• Achieved a 15x reduction in data processing costs by building RAG architecture with embedding and re-ranking models, accelerating drug discovery insights. • Enabled data extraction from scientific papers PDFs by implementing BERT-based document chunking. • Translated business goals into technical steps by coordinating with stakeholders on 5 targeted projects.
3	Data Scientist	At UBS Wealth Management, I build and deploy AI/**ML** models to strengthen portfolio risk management and optimize asset allocation. Using **XGBoost**, **LightGBM**, and statistical techniques, I developed predictive models and risk scoring algorithms that improved returns and reduced high-risk exposure. I collaborate with portfolio managers and analysts to translate complex data into actionable investment insights, while ensuring compliance with GDPR and SOX. I also design dashboards in Power BI and Tableau and maintain ETL pipelines in Python and SQL, enabling real-time decision-making supported by accurate, secure, and scalable data pipelines.
4	Vice President - AI/**ML** Product Manager - Portfolio Analysis Group	• Leading AI/**ML** product strategy for PAG, partnering with PAG analysts, front office, and risk teams to modernize exception management and performance attribution workflows. • Defined enterprise AI/**ML** roadmap for exception management and attribution workflows, aligning initiatives with Asset Management strategic priorities and presented results to senior PAG leadership. • Productionized a **random forest**-based after-tax outlier review solution, reducing workload by 2 FTEs. • Develop anomaly detection models for AM WRAP Composites, with the intention on reducing false positives. • Building NLP-based agent interactions to enable Front Office teams to query and derive insights from attribution datasets. • Oversee proof-of-concept initiatives for document extraction and composite description evaluation, scaling across PAG teams. • Conduct demos and training sessions for Front Office teams on new capabilities as well as gathering feedback.
5	Sr. Principal Systems Engineer	Working on advanced **neural network**s for radar and edge AI sensor processing. Working with external SMEs and Raytheon Fellows from across the country. Company AI/**ML** instructor. Codeveloped a tutorial on event-based sensing and **spiking neural network**s (SNNs).
6	Staff **Machine Learning** Engineer	Eater Incentive & Pricing team. I built the **ML**-based solutions to grant the right promotions/fees to the right users/cohorts. Applied the latest **ML** techniques such as sequential modeling, causal modeling et.al. to maximize the uplift of revenue and minimize the cost of promos, by giving personalized promotions to the consumers.
7	**Machine Learning** Engineer	Engineered an end-to-end ETL pipeline to process 52,000+ songs using the Spotify API, automating extraction and storage in PostgreSQL. Performed extensive **feature engineering** and data normalization for downstream **ML** applications Conducted a comprehensive **ML** study leveraging to disentangle and quantify the effects of audio attributes (e.g., tempo, danceability, acousticness) and metadata (e.g., genre, release year) on Spotify song popularity. **Pytorch** was used for building a ANN to model the relationship between 17 features. Incorporated model diagnostics, variance inflation factor analysis, and out-of-sample validation to ensure robustness; insights contributed to understanding content personalization and music recommendation strategies
8	**Machine Learning** Engineer	• Worked on team of 3 engineers to develop automated data collection pipelines for 14+ data sources using Python and SQL with Databricks platform; pipelines served as foundation for **ML** models to provide key marketing insights to downstream teams. • Spearheaded agile scrum meetings and serve as liason for team by successfully managing stakeholder communications, cross-functional collaboration, internal documentation, and vendor relations.
9	Product Manager	• Led end-to-end product development of an AI-driven embedded finance platform, enabling partner applications to deliver real-time credit decisions via APIs, reducing loan approval time from days to under 5 seconds. • Defined product vision, strategy, and roadmap aligned with business objectives; collaborated with engineering, data science, and compliance teams to ensure innovation within FCRA, ECOA, and GDPR frameworks. • Partnered with data scientists to design and deploy an explainable **ML**-based credit-scoring model, achieving 94% predictive accuracy and transparent outputs using SHAP-based interpretability. • Managed a backlog of 40+ features, prioritized initiatives using the RICE framework, and delivered the MVP 20% ahead of schedule while maintaining enterprise-grade reliability and uptime. • Launched a self-service partner dashboard enabling instant API integration, credit analytics, and performance monitoring - improving partner onboarding efficiency by 30%. • Established measurable success metrics and continuous feedback loops, tracking KPIs such as approval-rate lift (+15%), model drift, and partner retention to drive iterative product optimization.
10	Data Scientist	• Developed AI-driven image processing models for OLED TVs using **TensorFlow** and **PyTorch**, achieving a 25% improvement in image clarity (PSNR metric) and enabling real-time 4K upscaling for the Z95A series launch in September 2024. • Designed predictive maintenance models with **XGBoost** and Apache Spark for manufacturing equipment, reducing downtime by 20% (MTTR metric) and supporting sustainability goals showcased at CES 2025. • Built customer segmentation models using BERT and Snowflake for market re-entry, increasing targeted campaign ROI by 15% through enhanced audience precision and data-driven insights. • Implemented **MLOps** pipelines with **AWS SageMaker** and **MLflow**, automating model deployment for the "Panasonic Go" AI initiative, achieving 30% faster deployment cycles and 99.9% model uptime. • Created interactive Tableau dashboards for sustainability metrics, integrating Google BigQuery data, enabling 10% faster decision-making on carbon emission reductions, aligning with Panasonic's net-zero 2030 target.
Panel C: Natural Language Processing		

Continued on next page

Table A.12 continued

Example	Job title	Job description
1	Data Engineer	Developed a predictive model using CatBoost to forecast credit risk, achieving a 15% improvement in prediction accuracy (Mean Absolute Error) over prior models. This refinement has significantly strengthened our risk management capabilities. Spearheaded research using Bertopic and LDAvis for **NLP topic modeling**, effectively pinpointing key themes in customer feedback across various channels. This initiative has directly influenced product improvements and customer service strategies. Created and fine-tuned an ARIMAX model to accurately predict monthly loan demand, directly contributing to $590,000 in annual savings through optimized resource allocation and enhanced financial planning.
2	Business Data Analyst	Built and automated 35+ dashboards using Tableau, Qlik, and Streamlit, reducing manual reporting time by 60% across departments. Analyzed structured & semi-structured data with Python and SQL, boosting campaign ROI and cutting churn by 25%. Optimized ETL pipelines with SSIS and Qlik Data Integration, improving refresh speed by 70% and reducing latency. Designed scalable data models and star schemas across Snowflake, MySQL, and Qlik, enabling real-time analytics for 5+ teams. Cleansed and transformed datasets with 10M+ rows, ensuring accurate reporting for finance, sales, and operations. Analyzed 100K+ customer reviews, increasing **sentiment classification** precision by 18% and informing product strategy. Created Excel models with pivot tables, macros, and lookups, cutting spreadsheet errors by 35% and enabling faster financial scenarios. Developed interactive Qlik Sense dashboards with drill-downs, improving decision-making speed by 40%. Automated data ingestion from 10+ sources using Qlik pipelines, reducing manual prep by 50%. Enhanced dashboard performance with advanced Qlik scripting (SET analysis, variables), cutting load times by 30% on 20M+ row datasets
3	Software Engineer	• Developed an AI Assist Intelligent Bot using Oracle cloud, Python, and **NLP**, and trained models for context-sensitive user assistance to enable real-time information retrieval and interactive support, significantly enhancing user engagement and satisfaction. • Developed an interactive dashboard for business users to provide better insights into metrics, KPIs, and trends on a SaaS application. Utilized Python, HTML, CSS, and JavaScript, following an API-first approach microservice-based architecture, and Integrated it with the AI Help Tool. • Served as Scrum Master, effectively managing sprint planning and execution using Jira. Created and prioritized user stories, logged work, and facilitated agile ceremonies to improve team productivity and Project outcomes. • Enhanced my knowledge in Oracle SaaS and AI, earning key certifications and applying these skills to drive innovative solutions and business success.
4	**NLP** Algorithm Engineer	Worked on text processing projects, focusing on **Named Entity Recognition** (NER) related to e-commerce, leveraging **Natural Language Processing** (**NLP**) techniques to enhance the recommendation system's accuracy and relevance. Collaborated closely with cross-functional teams to identify and resolve complex challenges, ensuring seamless project execution.
5	Staff Software Engineer (Apple Intelligence)	Scalable Knowledge Graphs for On-Device Intelligence • Designed and implemented scalable knowledge graph architectures optimized for inference, enabling real-time contextual intelligence in resource-constrained environments. • Developed graph ingestion pipelines to process large-scale heterogeneous data sources, ensuring efficient entity resolution, **relationship extraction**, and semantic enrichment. • Integrated lightweight graph query engines and custom embeddings to support intelligent decision-making on edge devices. • Focused on optimization for latency, memory footprint, and incremental updates to enable robust offline performance and personalization at scale. • On-device embeddings computation and store it in Vector DB
6	Data Scientist	Developed and trained a **natural language processing** (**NLP**) model to automate invoice processing, boosting the rate of fully automated invoices from 13% to 60% and significantly reducing manual effort and processing costs.
7	Senior Consultant	Specialized in forensic analytics with hands-on experience supporting Anti-Money Laundering (AML) and Bank Secrecy Act (BSA) compliance investigations. Designed and implemented ETL pipelines, developed automation solutions, and built predictive models leveraging advanced **natural language processing** (**NLP**) techniques.
8	AI Software Developer	**NLP**-driven land price forecasting.
9	Data Scientist	• Developed and deployed predictive models, enhancing customer engagement and informing data-driven strategies. • Engineered ETL pipelines and clustering solutions for effective customer segmentation, optimizing marketing efforts. • Created executive Tableau dashboards and utilized **NLP** for **sentiment analysis**, providing actionable insights for leadership.
10	Data Analyst	Developed scalable data workflows using Python and SQL to streamline data collection, wrangling, and preprocessing tasks, enabling timely delivery of analytical insights to business stakeholders. Applied statistical modeling and predictive analytics techniques to support 20+ business initiatives, improving operational forecasting accuracy by 30% and enabling proactive resource planning. Designed and launched 15+ interactive dashboards using Tableau and Excel, increasing executive reporting engagement by 40% through effective data storytelling and visualization strategies. Developed and maintained custom analytics models in Python (Pandas, Scikit-learn), identifying key behavioral patterns that led to a 25% boost in customer retention and a 20% increase in sales conversion. Built and automated cloud-based data pipelines across AWS and GCP environments, reducing manual intervention by 60% and achieving 99.9% uptime for real-time analytics processing. Constructed and validated supervised and unsupervised ML models, including classification, regression, clustering, and **NLP** models, leading to a 40% increase in predictive accuracy for business-critical use cases. Partnered with business and technical teams to document data requirements, deploy version-controlled code via GitHub, and ensure collaboration throughout the project lifecycle.
Panel D: Computer Vision		

Continued on next page

Table A.12 continued

Example	Job title	Job description
1	Data Scientist	Developed and deployed classical **computer vision** algorithms (e.g., image filtering, morphological operations, feature extraction, **template matching**, object segmentation) to automate manufacturing processes in aluminum recycling, increasing system efficiency and reducing manual intervention Engineered performance metrics and designed key performance indicators (KPIs) to assess and enhance the accuracy of feature detection, resulting in a 10% improvement in detection precision Implemented a multi-threaded image processing pipeline, optimizing real-time feature detection performance, reducing processing latency by 60% Calibrated machine vision cameras to ensure precise pixel-to-world distance measurements and consistent object localization Collaborated with the **computer vision** team to integrate new algorithms into the existing pipeline, utilizing continuous integration workflows via GitHub Projects to streamline development and ensure seamless version control and testing
2	AI Research Scientist	Development of a Generalized Pipeline for Carotid Vessel Segmentation Across Diverse MRI Sequences: Advanced Research in Data Processing, Normalization, and **Computer Vision** Models Optimization Leveraging Supercomputing.
3	Senior Data Analyst	• Analyze large volumes of supply chain data to identify bottlenecks, track KPIs, and support strategic decision-making in reverse logistics. • Collaborate with cross-functional teams to integrate AI tools into operational systems, achieving a 20% improvement in processing efficiency. • Led AI-driven process improvements in Verizon's reverse logistics operations, using **computer vision** (YOLO, CNN) to optimize grading machine performance and increase accuracy of returned product assessments. • Apply machine learning models to automate and enhance refund approval workflows, reducing processing time and human error.
4	Senior **Computer Vision** Scientist	Working on the Research and Development of framework for AI assisted eye surgery
5	Software Engineer, Fundamental AI Research (FAIR)	Core contributor to the Segment Anything 2 (SAM 2) research demo, which served the SAM 2 model and all its capabilities to x00,000 global users. Building new research playgrounds, human annotation tasks, and data visualization tools for **computer vision** research (object segmentation in video, semantic grounding in images and videos) FAIR (Fundamental AI Research) - Perception
6	Senior Software Engineer	The company developed concentrated solar power (CSP) technology including novel heliostats and multithreaded C# control software • Created the heliostat battery management system's central controller, which used MQTT protocol; to prevent total battery discharge, functions included a low power mode, a nightly power off, and a failsafe heartbeat; enabled a 21% increase in test site uptime • Improved a **computer vision** gradient anomaly mask for cloud detection and created a test visualization tool to enable future work • Performed a root cause analysis to successfully identify temporary server overloads as the cause of communication losses • Collaborated on robustness improvements to the 4th generation heliostats' TCP/IP-to-serial network bridge and protocol handler
7	Data Scientist	**Computer Vision** (AI driven content for the website and identifying products in images)
8	Principal Solutions Architect	As Principal Solutions Architect at Berkshire Grey, I lead the end-to-end design and deployment of AI-powered robotic automation systems that optimize large-scale supply chain operations. My role blends deep technical architecture with strategic leadership, guiding cross-functional teams to deliver complex solutions that integrate robotics, AI/ML, **computer vision**, and orchestration software. I work directly with enterprise clients to translate operational challenges into scalable automation strategies, while also influencing product direction through field-driven insights. Focused on execution and impact, I've delivered high-throughput systems that drive measurable efficiency gains across retail, eCommerce, and logistics environments.
9	Director of Engineering	I lead and manage multiple **computer vision**/AI/machine learning engineering teams, developing organizational-level technology strategies and roadmaps to ensure engineering deliverables align with Intel's overall business and product strategy.
10	AP Engineer	. Develop and optimize AI models for autonomous driving, robotics, and energy systems, focusing on scalability and real-world application.. Design deep learning algorithms for **computer vision**, sensor fusion, and predictive analytics to enhance Tesla's Full Self-Driving (FSD) capabilities. • Process and analyze large-scale datasets from Tesla's fleet to identify and address edge cases.. Collaborate with cross-functional teams to integrate AI solutions into production and proprietary hardware.. Optimize machine learning models for deployment on Tesla's FSD hardware, ensuring minimal latency and high performance.. Conduct research on emerging AI technologies to drive innovation aligned with Tesla's mission of accelerating sustainable energy adoption.
Panel E: Generative AI		
1	Data Science / AI Engineer	• Implementing best-practice **LLM** prompting (prompt chaining and modularization) for concurrent taxonomy generation and content tagging to improve B.O. prediction • Utilizing agent-based prompting (task agent and verification agent) to generate in-depth feature explanations for multidimensional film similarity searches, ensuring **LLM** answers are grounded in real tabular data. • Proposing and testing various clustering, representation similarity, and A/B testing strategies to validate attributes generated by **LLMs**
2	Applied Scientist	**Large language models** for Amazon shopping and customer understanding. Pre-training (long context and hybrid attention architectures). Pre-train evaluations.
3	Full-stack Developer	Developed a **GenAI**-powered customer insights engine integrating GPT APIs & LangChain with Spring Boot microservices. Built RAG pipelines (**Pinecone**, FAISS) enabling semantic search across financial documents & compliance records. Designed AI-powered chatbots & knowledge assistants (React.js + Spring Boot), improving support accuracy by 40%. Integrated Kafka event streaming + **LLM** APIs for real-time fraud detection & anomaly insights. Deployed **LLMs** on AWS (Bedrock, SageMaker) for scalable, hybrid rule-based + AI decision systems. Implemented secure **prompt engineering**, API throttling, and logging for AI observability (Splunk, Grafana). Enhanced backend for transaction monitoring & fraud detection while embedding **GenAI** features into workflows.

Continued on next page

Table A.12 continued

Example	Job title	Job description
4	**GenAI** Engineer	I design and deploy **LLM**-powered **Generative AI** applications using LangChain, Hugging Face, and OpenAI APIs, integrating RAG pipelines for enterprise-scale knowledge retrieval. I have fine-tuned multilingual models like T5 and GPT-Neo with DeepSpeed and Accelerate, boosting domain-specific accuracy by up to 20%. My work includes building document Q&A systems, AI chatbots, and multi-agent frameworks with FAISS and **Pinecone**, and implementing scalable inference via Triton. I also design benchmarking dashboards to track latency, hallucination, and token usage, and lead A/B testing across **GPT-4**, Claude, and Cohere, optimizing performance and reliability.
5	**Generative AI** Engineer	Designing and deploying enterprise-grade **Generative AI** solutions with a focus on **LLMs**, RAG pipelines, and agentic AI systems. Built autonomous AI workflows (AutoGen, CrewAI) to automate reasoning and multi-step tasks, integrated with Azure + **vector databases** (**Pinecone**). Deployed scalable models via Docker, Kubernetes, REST APIs, and implemented monitoring with Grafana & Prometheus. Applied evaluation frameworks (RAGAS, G-Eval) to ensure accuracy, safety, and business relevance.
6	Senior Dot Net Full Stack Developer	Architecting operational data mesh on ADLS Gen2, Delta Lake, Unity Catalog Implementing PySpark ETL frameworks with schema governance, incremental upserts Integrating **Azure OpenAI** embeddings & RAG pipelines into operational datasets Designing CDC ingestion using Event Hubs + Delta checkpoints Building FastAPI microservices exposing reliability signals Creating ADF pipelines with lineage and dependency graphs Developing reliability semantic layers in Looker for SLO/SLA reporting Instrumenting ETL & **LLM** workflows using Azure Monitor, App Insights Enforcing cost governance & secure operational boundaries Automating CI/CD using GitHub Actions for notebooks and Delta tables
7	Senior Software Engineer	Designed and deployed end-to-end Voice AI Agents using LiveKit, Retell AI, Vapi, and Twilio, integrating real-time STT, LLM_x0002_based dialogue management, and natural TTS responses to handle dynamic phone conversations, automate customer service workflows, and connect seamlessly with CRM and backend APIs through n8n and webhooks for data synchronization and task automation.
8	AI Enablement & Community Adoption	At AT&T, I've had the privilege of leading and contributing to several transformative AI initiatives-most notably the design, development, and deployment of advanced AI assistants powered by **Retrieval-Augmented Generation** (RAG) and fine-tuned **large language models** (**LLMs**). One of the most exciting projects I spearheaded was the creation of a multi-purpose RAG-based AI assistant that supports customer service, sales, and technical troubleshooting teams. This assistant draws from AT&T's vast ecosystem of structured and unstructured data-ranging from product catalogs and billing documentation to network troubleshooting manuals and customer interaction histories-enabling it to deliver highly accurate, context-aware, and grounded responses in real-time. To build this, I integrated **LLMs** like OpenAI's GPT, Anthropic's Claude, Cohere, and Mistral with LangChain for orchestration and **Pinecone**/FAISS for semantic retrieval. By combining the power of foundation models with precise information retrieval systems, we created a tool that drastically reduced time-to-resolution for customers and improved operational efficiency for internal teams. But generic **LLMs** don't understand domain-specific language or nuances-so I fine-tuned models like Mistral and LLaMA on AT&T-specific datasets including customer service transcripts, FAQs, and internal technical documentation. This improved accuracy, increased trust in AI responses, and allowed us to maintain relevance even in highly specialized telecom conversations. I also developed optimized **GenAI** pipelines involving embedding refinement, dynamic chunking strategies, and multi-step prompt chaining. These pipelines ensured our assistants could not only retrieve the right knowledge but synthesize it effectively-even when customers asked layered, ambiguous, or multi-intent queries. The result was a robust conversational system capable of delivering accurate, contextual, and natural-sounding answers that mirrored human-level support.
9	Associate Engineer, AI	- AI & Advanced Analytics- Internal **Generative AI** use-case/experimentation enablement
10	Analyst	Product Lead on SnackGPT, a **GenAI** platform that analyzed complex sets of business data using Gemini and YAML query engineering. Responsibilities involved leading daily stand-up meetings with business and tech employees, testing the app, and resolving front- and back-end bugs when they arose. Front-end developer: Designed, built, tested and deployed several websites and dashboards that were used in day-to-day business activities. Products included the NA Digital Portal, a unified landing page built using JavaScript integrated with Metric Insights API, and several Power BI dashboards. Business products involved querying and transforming large sets of data from sources like Google Cloud Platform and designing a convenient UX, as well as leading training sessions with 200+ business users to introduce the product.
Panel F: Agentic AI		
1	Software Engineer	Implemented a RAG chatbot using Hugging Face, **LangChain**, and Oracle Autonomous Database to enable semantic search across internal docs, adopted by internal engineering teams.
2	Head of Product, Agentforce Guardrails	Building AgentForce - the world's first and most innovative platform for Trusted **AI Agents** for enterprise. I lead the product vertical for trusted, safe and secure **AI Agents**. Learn more:
3	Sr Manager, Product Management - **Agentic AI**	• Defined and led WEX's enterprise **Agentic AI** strategy ($2B+ portfolio) in partnership with CDO, CTO, and business leaders across Mobility, Corporate Payments, and Benefits. • Built consensus among 15+ stakeholders across Technology, Operations, and Customer Experience to define enterprise Virtual Assistant strategy, architecting unified multi-agent ecosystem (Benefits Assist, Claims Processing AI, Agent Assist) with human-centered design principles and API-first microservices architecture deployed across voice, web, messaging, and mobile channels. • Scaled Virtual Assistant platform to deliver measurable business impact powering 5M+ annual interactions, improving self-service containment, reducing average handle time, and generating $2M+ in operational savings while maintaining 4.4/5 customer satisfaction score. • Spearheaded Finn, an enterprise agentic marketplace unlocking $2M(25K hours) in productivity savings. • Managed PMs (ICs and Managers) and a team of 80+ engineers, shipping embedded GenAI products (AutoIVR, Claims Automation, Fleet Intelligence) generating $10M+ in new revenue. • Built AI partner ecosystem through strategic partnerships with UiPath (RPA), ThoughtSpot, Moveworks, Decagon, AWS, GCP and OpenAI, accelerating intelligent automation, and enabling scalable GenAI enterprise adoption. • Oversaw 40+ AI experiments across diverse customer segments to validate pain points and de-risk product ideas.

Continued on next page

Table A.12 continued

Example	Job title	Job description
4	Gen AI Developer	Engineered the AI Lending Insight Platform leveraging LLaMA-based generative AI for contextual credit and payment behavior analysis, generating narrative explanations for loan approval, BNPL eligibility, and transaction patterns, reducing manual review effort by ~12%. • Integrated generative AI into Python-based backend microservices (FastAPI/Flask), embedding LLaMA inference to generate real-time consumer credit summaries and payment recommendations, accelerating decision-making turnaround by nearly 15% for lending and risk teams. • Orchestrated **AI agents** to correlate transactional behavior, repayment history, and alternative financial signals while designing and optimizing PostgreSQL and Aurora schemas, applying indexing, partitioning, and consistency strategies to support high-volume consumer lending workloads. • Architected a BNPL risk and payment decision microservices using Python, Kafka, and gRPC, using OAuth authentication, AWS Cognito, end-to-end encryption, secrets management, JWT validation, audit logging, and RBAC, ensuring platform security and regulatory compliance. • Developed real-time merchant and consumer dashboards using React, TypeScript, Redux Toolkit, WebSockets, and D3.js to visualize payment trends, credit utilization, and risk indicators. Applied unit and integration testing with Pytest, Selenium, and automated CI/CD pipelines, improving release reliability by 13%. • Collaborated with DevOps teams to deploy containerized Python services on AWS EKS using Docker and Kubernetes, integrating OpenTelemetry, Prometheus, Grafana, and ELK for enhanced observability, autoscaling, and production stability, improving system uptime and monitoring efficiency.
5	Software Engineer II	Developing tools to objectively measure and refine the performance of **AI agents**. Ensuring every AI-driven assistant stays aligned with your goals and keeps improving through guided feedback.
6	Backend & Gen AI Engineer	• Designed, built, and scaled a production-grade **Agentic AI** platform delivering conversational assistants, task-based copilots, and autonomous **AI agents** for enterprise analytics and decision support. • Evaluated and integrated LLMs and multimodal foundation models from multiple providers (OpenAI, Google, Anthropic, open-source via Hugging Face/Ollama), selecting models dynamically based on latency, cost, and task complexity. • Architected agentic workflows using **LangChain** and OpenAI Agents SDK, enabling tool-calling, multi-step reasoning, memory, and multi-agent orchestration to autonomously plan and execute complex tasks. • Built advanced RAG pipelines with: • Vector embeddings and databases (Pinecone / PGVector / Vertex AI Search) • Multi-source retrieval across documents, tickets, logs, and structured data • Reranking strategies and semantic caching to improve relevance and response quality • Implemented prompt engineering best practices, including system instructions, few-shot prompting, JSON/JSON Schema structured outputs, and function/tool calling to ensure deterministic and auditable AI behavior. • Integrated safety and guardrail layers (content moderation APIs, custom policies, schema validation, hallucination detection) to minimize unsafe outputs and keep conversations focused and compliant. • Developed full-stack AI experiences, integrating **AI agents** into React-based front-end applications with real-time interactions via WebSockets and secure backend orchestration. • Built scalable backend services using Python, exposing AI orchestration via REST, gRPC, and WebSockets, and deployed them in containerized, cloud-native environments using Docker and Kubernetes. • Designed and maintained LLM evaluation and monitoring pipelines, tracking offline and online metrics such as accuracy, relevance, latency, and cost; used LangSmith and OpenTelemetry for tracing and observability.
7	Senior Engineering Manager - Virtual Expert Platform	• Innovate in the assisted services domain using **agentic ai** to build done for you experiences. • Lead engineers on a few core engineering teams to enhance the Virtual Expert Platform, driving customer-expert engagements. • Implement AI-driven workflows that streamline interactions across Intuit products, creating seamless user experiences. • Foster collaboration among diverse personas to develop innovative agentic solutions, significantly improving customer satisfaction.
8	Python Developer	Designed and implemented FastAPI-based microservices for real-time trade risk analytics, reducing API latency by 42%. • Developed automated ETL pipelines using Python, Pandas, and SQLAlchemy to sync trade data between internal systems and third party platforms (e.g., Bloomberg). • Integrated JWT/OAuth2-secured APIs for account and portfolio services, ensuring robust authorization and compliance with FINRA and SOX guidelines. • Containerized applications using Docker and orchestrated deployments via Kubernetes (EKS) for high availability and scalability. • Optimized PostgreSQL queries and database indexing strategies, resulting in a 35% performance gain for analytical dashboards. • Created internal Python libraries for financial data transformation, utilized across 5+ teams for standardization. • Integrated LLM-based summarization models (using OpenAI APIs and **LangChain**) for generating investment insights and market summaries. • Collaborated in Agile teams with product managers, DevOps, and data engineers using Jira, Confluence, and GitHub Actions

Continued on next page

Table A.12 continued

Example	Job title	Job description
9	Program Manager	• Spearheaded GenAI business review cadences (5 VP-level MBRs), aligning Directors across 25+ initiatives and directly reprioritizing $15M+ investments in **Agentic AI** and Foundational Tech programs • Orchestrated 7 cross-functional (Product, Engineering and Legal) partnerships to streamline GenAI experimentation pipelines, accelerating time-to-feedback by 45% and enabling 3 high-impact pilots to scale across customer touchpoints • Drove 50% increase in GenAI program adoption by launching an impact-tracking system that influenced $10M+ in funding decisions and helped sunset 3 low-ROI initiatives unlocking $5M+ in variable cost efficiency • Orchestrated cross-portfolio revenue forecasting models through BIE partnerships spanning 25+ initiatives, enabling VP-level strategic pivots that redirected $15M in GenAI investments toward highest-ROI opportunities • Deployed scalable automation dashboards covering 120+ tech teams, reducing manual reporting overhead by 70% and informing 3 engineering practice pivots based on surfaced telemetry patterns • Directed automated pipeline attribution system development across 120+ product teams, surfacing performance correlation patterns that informed 3 strategic operational pivots and improved delivery timeline accuracy by 35%
10	Lead Generative AI Engineer	Enterprise AI-Powered Clinical & Claims Automation Platform • Architected enterprise-grade LLM-powered Clinical Decision Support System using GPT-4, **LangChain**, and RAG 2.0 • Designed scalable **Agentic AI** frameworks with multi-agent orchestration and tool calling • Built AI microservices using Python (FastAPI) and Java (Spring Boot) • Implemented HIPAA-compliant AWS deployments (EKS, Lambda, S3, IAM) • Integrated AI modules with legacy J2EE healthcare systems and HL7/FHIR APIs • Developed RAG pipelines using LlamaIndex, Pinecone, FAISS • Implemented Kafka-driven real-time claims processing • Led LLM evaluation frameworks with A/B testing and human-in-the-loop validation • Ensured enterprise security (OAuth2, JWT, RBAC, AES-256 encryption)

This table reports 10 examples of AI-related U.S. jobs by AI subcategory for the U.S. public-firm sample from 2010 to 2024. Jobs are assigned to subcategories based on keyword matches in either the job title or the job description. The job descriptions are shown to illustrate the type of text captured by the keyword-based classification, and bold text highlights AI-related terms or phrases under each subcategory.

Table A.13. Top 30 Job Titles with the Most AI Jobs in 2024

Rank	Job-title	AI positions	Rank	Job-title	AI positions
1	Software Engineer	19,515	16	Director of Product Management	725
2	Machine Learning Engineer	9,792	17	Machine Learning Scientist / Researcher	665
3	Data Scientist	8,170	18	Product Designer	637
4	Product Manager	4,695	19	Account Executive	630
5	Data Engineer	4,523	20	Business Analyst	548
6	Data Analyst	4,120	21	Vice President	413
7	Engineering Manager	3,974	22	Manager	376
8	AI Engineer	3,778	23	AI Software Engineer	312
9	Program / Project Manager	2,555	24	Data Science Manager	307
10	Research Scientist	1,900	25	Solutions Architect	269
11	Software Development Engineer	1,411	26	Technical Recruiter	242
12	Software Developer	1,406	27	Computer Vision Engineer	203
13	Director	1,373	28	Generative AI Associate	183
14	Applied Scientist	1,127	29	Artificial Intelligence Partner	179
15	Technical Lead / Principal Engineer	1,106	30	Product Marketing Manager	168

This table reports the 30 job titles with the largest number of AI-related U.S. jobs in 2024 for the U.S. public-firm sample. Job titles are ranked by the total number of AI-related jobs. Titles are based on raw LinkedIn job-title strings observed in Revelio. To generate this table, we aggregate closely related title variants ourselves: seniority and level modifiers, such as "Senior," "Staff," "Principal," "Lead," "II," and "III," are combined when the underlying role is the same. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.14. Top 10 Job Titles with the Most AI Jobs by AI Subcategory in 2024

Rank	AI	ML	NLP	CV	Gen AI	Agentic AI
1	Software Engineer	Machine Learning Engineer	Data Scientist	Software Engineer	Software Engineer	Software Engineer
2	Product Manager	Software Engineer	Software Engineer	Data Scientist	Data Scientist	Data Scientist
3	Data Scientist	Data Scientist	Machine Learning Engineer	Computer Vision Engineer	Machine Learning Engineer	AI/ML Engineer
4	AI Engineer	Data Engineer	AI/ML Engineer	Machine Learning Engineer	Product Manager	Machine Learning Engineer
5	Machine Learning Engineer	Data Analyst	Data Engineer	AI Engineer	Applied Scientist	Product Manager
6	AI/ML Engineer	AI/ML Engineer	Data Analyst	AI/ML Engineer	Data Engineer	AI Engineer
7	Data Engineer	Product Manager	AI Engineer	Artificial Intelligence Engineer	Software Development Engineer	Data Engineer
8	Technical Program Manager	Machine Learning Software Engineer	Applied Scientist	Applied Scientist	AI/ML Engineer	Generative AI Engineer
9	Engineering Manager	Python Developer	Product Manager	Research Scientist	Research Scientist	Software Development Engineer
10	Program Manager	Engineering Manager	Artificial Intelligence Engineer	Autonomous Vehicles Software Engineering Manager	AI Engineer	Gen AI Engineer

This table reports the 10 job titles with the largest number of AI-related U.S. jobs in 2024 within each AI subcategory. Job titles are ranked by the number of AI-related jobs within each subcategory. Titles are based on raw LinkedIn job-title strings observed in Revelio. To generate this table, we aggregate closely related title variants ourselves: seniority and level modifiers, such as "Senior," "Staff," "Principal," "Lead," "II," and "III," are combined when the underlying role is the same. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.15. Top 30 Occupations with the Highest AI Shares in 2024

Rank	O*NET code	Occupation	AI share	AI jobs	Total jobs
1	15-1221.00	Computer and Information Research Scientists	0.66	23,985	36,568
2	15-2051.00	Data Scientists	0.43	50,403	117,157
3	15-2099.00	Mathematical Science Occupations, All Other	0.39	335	867
4	17-2199.08	Robotics Engineers	0.24	1,515	6,222
5	25-1022.00	Mathematical Science Teachers, Postsecondary	0.22	68	305
6	19-2099.01	Remote Sensing Scientists and Technologists	0.17	46	265
7	19-1029.01	Bioinformatics Scientists	0.17	785	4,611
8	15-1255.01	Video Game Designers	0.14	532	3,846
9	19-3022.00	Survey Researchers	0.14	999	7,363
10	15-1299.07	Blockchain Engineers	0.13	1,605	12,313
11	15-1243.00	Database Architects	0.11	12,253	107,464
12	29-2036.00	Medical Dosimetrists	0.10	69	676
13	15-1255.00	Web and Digital Interface Designers	0.10	6,110	61,608
14	27-3091.00	Interpreters and Translators	0.10	227	2,353
15	25-1021.00	Computer Science Teachers, Postsecondary	0.09	180	1,971
16	17-2031.00	Bioengineers and Biomedical Engineers	0.08	1,364	16,404
17	13-2099.01	Financial Quantitative Analysts	0.08	1,103	13,713
18	15-2021.00	Mathematicians	0.08	42	526
19	19-3041.00	Sociologists	0.08	25	314
20	13-1161.01	Search Marketing Strategists	0.06	2,026	32,151
21	19-2012.00	Physicists	0.06	36	581
22	17-2072.01	Radio Frequency Identification Device Specialists	0.06	1,959	31,756
23	25-1042.00	Biological Science Teachers, Postsecondary	0.06	26	432
24	17-2199.06	Microsystems Engineers	0.06	8,247	139,429
25	19-1099.00	Life Scientists, All Other	0.06	425	7,369
26	27-1029.00	Designers, All Other	0.06	2,365	42,656
27	15-1252.00	Software Developers	0.06	40,418	729,588
28	19-3094.00	Political Scientists	0.06	62	1,125
29	19-3011.00	Economists	0.05	118	2,347
30	11-2021.00	Marketing Managers	0.05	17,849	358,050

This table reports the 30 occupations with the highest AI-related job shares among U.S. jobs in 2024 for the U.S. public-firm sample. Occupations are based on O*NET occupation codes. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs in the same occupation and is reported as a fraction. The sample is restricted to firm-years with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to occupations with at least 100 total U.S. jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.16. Top 30 Occupations with the Most AI Jobs in 2024

Rank	O*NET code	Occupation	AI jobs	AI share	Total jobs
1	15-2051.00	Data Scientists	50,403	0.4302	117,157
2	15-1252.00	Software Developers	40,418	0.0554	729,588
3	15-1221.00	Computer and Information Research Scientists	23,985	0.6559	36,568
4	11-2021.00	Marketing Managers	17,849	0.0499	358,050
5	15-1299.09	Information Technology Project Managers	16,234	0.0365	444,979
6	15-2051.01	Business Intelligence Analysts	14,719	0.0456	323,089
7	15-1299.08	Computer Systems Engineers/Architects	12,667	0.0403	314,006
8	15-1243.00	Database Architects	12,253	0.114	107,464
9	17-2199.06	Microsystems Engineers	8,247	0.0591	139,429
10	15-1255.00	Web and Digital Interface Designers	6,110	0.0992	61,608
11	41-9031.00	Sales Engineers	5,126	0.0432	118,686
12	13-1071.00	Human Resources Specialists	4,196	0.027	155,460
13	41-3091.00	Sales Representatives of Services, Except Advertising, Insurance, Financial Services, and Travel	3,413	0.0152	224,262
14	11-1011.00	Chief Executives	3,395	0.0187	181,871
15	17-2199.00	Engineers, All Other	3,092	0.0171	180,902
16	15-1299.05	Information Security Engineers	2,713	0.0371	73,218
17	11-2022.00	Sales Managers	2,618	0.0081	324,345
18	19-1029.00	Biological Scientists, All Other	2,586	0.0294	87,953
19	11-3021.00	Computer and Information Systems Managers	2,398	0.0315	76,132
20	27-1029.00	Designers, All Other	2,365	0.0554	42,656
21	23-1011.00	Lawyers	2,336	0.0326	71,718
22	41-4012.00	Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products	2,214	0.006	367,741
23	15-1254.00	Web Developers	2,213	0.0285	77,710
24	13-1161.00	Market Research Analysts and Marketing Specialists	2,171	0.0493	44,005
25	11-3031.01	Treasurers and Controllers	2,070	0.0074	279,856
26	13-1199.00	Business Operations Specialists, All Other	2,039	0.0168	121,426
27	13-1161.01	Search Marketing Strategists	2,026	0.063	32,151
28	17-2072.01	Radio Frequency Identification Device Specialists	1,959	0.0617	31,756
29	15-1253.00	Software Quality Assurance Analysts and Testers	1,949	0.0175	111,525
30	15-1243.01	Data Warehousing Specialists	1,912	0.0319	59,908

This table reports the 30 occupations with the largest number of AI-related U.S. jobs in 2024 for the U.S. public-firm sample. Occupations are based on O*NET occupation codes. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs in the same occupation and is reported as a fraction. The sample is restricted to firm-years with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to occupations with at least 100 total U.S. jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.17. Top 3 Firms with Highest AI Shares by Industry in 2024

Industry	Rank	Firm	AI share	AI jobs	Total jobs
Accommodation & Food Svcs	1	SONDER HOLDINGS INC	0.0118	12	1,015
	2	MGM RESORTS INTERNATIONAL	0.0073	72	9,809
	3	BOYD GAMING CORP	0.0070	15	2,145
Admin Support	1	STEM INC	0.1308	62	474
	2	KNIGHTSCOPE INC	0.1061	14	132
	3	RESEARCH SOLUTIONS INC	0.0800	6	75
Arts/Entertainment	1	DRAFTKINGS INC	0.0177	75	4,243
	2	FLUTTER ENTERTAINMENT PLC	0.0172	1	58
	3	ENDEAVOR GROUP HOLDINGS	0.0091	22	2,429
Construction	1	ORION GROUP HOLDINGS INC	0.0122	3	246
	2	BEAZER HOMES USA INC	0.0077	12	1,556
	3	FERROVIAL SE	0.0068	5	731
Education Svcs	1	NERDY INC	0.1835	29	158
	2	TAL EDUCATION GROUP	0.1194	8	67
	3	DUOLINGO INC	0.0749	100	1,335
Finance/Insurance	1	UPSTART HOLDINGS INC	0.1025	199	1,941
	2	DAVE INC	0.0687	29	422
	3	VISA INC	0.0677	1,316	19,450
Health Care	1	TEMPUS AI INC	0.1134	403	3,555
	2	DARIOHEALTH CORP	0.0793	18	227
	3	ONTRAK INC	0.0633	15	237
Information	1	C3.AI INC	0.5156	960	1,862
	2	COVEO SOLUTIONS INC	0.3553	27	76
	3	BAIDU INC	0.3434	114	332
Manufacturing	1	XPENG INC	0.3645	109	299
	2	NVIDIA CORP	0.3007	8,965	29,818
	3	ABSCI CORP	0.2301	55	239
Mining	1	ENCORE ENERGY CORP	0.0192	1	52
	2	SLB LTD	0.0177	302	17,080
	3	VITAL ENGY INC	0.0176	9	511
Other Svcs	1	FRONTDOOR INC	0.0109	16	1,472
	2	WW INTERNATIONAL INC	0.0109	39	3,590
	3	CHARGEPOINT HOLDINGS INC	0.0094	11	1,166
Prof & Business Svcs	1	INNODATA INC	0.7169	595	830
	2	STAGWELL INC	0.1193	109	914
	3	FLUENT INC	0.1066	37	347
Real Estate	1	CEVA INC	0.1613	10	62
	2	INTERDIGITAL INC	0.0613	20	326
	3	DOLBY LABORATORIES INC	0.0466	77	1,651
Retail Trade	1	COUPANG INC	0.2098	214	1,020
	2	ALIBABA GROUP HLDG	0.1197	116	969
	3	XOMETRY INC	0.0702	61	869
Transportation/Warehousing	1	DIDI GLOBAL INC	0.1079	26	241
	2	SERVE ROBOTICS INC	0.1061	21	198
	3	GRAB HOLDINGS LIMITED	0.0394	5	127
Utilities	1	NEXTDECADE CORP	0.0211	6	284
	2	NEXTERA ENERGY INC	0.0195	191	9,801
	3	AES CORP (THE)	0.0185	70	3,791
Wholesale	1	INSIGHT ENTERPRISES INC	0.0361	318	8,812
	2	EPLUS INC	0.0285	58	2,037
	3	MCKESSON CORP	0.0266	574	21,541

This table reports the three firms with the highest AI-related job shares within each sector (based on 2-digit NAICS industry codes) in 2024 for the U.S. public-firm sample. Rank is assigned separately within each industry. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.18. Top 3 Firms with Most AI Jobs by Industry in 2024

Industry	Rank	Firm	AI jobs	AI share	Total jobs
Accommodation & Food Svcs	1	STARBUCKS CORP	253	0.0021	119,293
	2	MARRIOTT INTL INC	159	0.0036	44,168
	3	MCDONALD'S CORP	146	0.0023	62,615
Admin Support	1	EXPEDIA GROUP INC	801	0.0670	11,960
	2	S&P GLOBAL INC	457	0.0312	14,637
	3	EQUIFAX INC	381	0.0401	9,504
Arts/Entertainment	1	DRAFTKINGS INC	75	0.0177	4,243
	2	LIVE NATION ENTERTAINMENT	73	0.0047	15,632
	3	BRIGHTSTAR LOTTERY PLC	31	0.0060	5,166
Construction	1	LENNAR CORP	55	0.0058	9,483
	2	FLUOR CORP	19	0.0018	10,346
	3	BEAZER HOMES USA INC	12	0.0077	1,556
Education Svcs	1	PEARSON PLC	312	0.0220	14,212
	2	DUOLINGO INC	100	0.0749	1,335
	3	SKILLSOFT CORP	96	0.0620	1,548
Finance/Insurance	1	JPMORGAN CHASE & CO	4,527	0.0236	191,930
	2	CAPITAL ONE FINANCIAL CORP	3,284	0.0458	71,757
	3	WELLS FARGO & CO	2,699	0.0130	206,854
Health Care	1	CVS HEALTH CORP	2,173	0.0165	131,826
	2	HUMANA INC	910	0.0228	39,905
	3	HCA HEALTHCARE INC	417	0.0089	46,674
Information	1	MICROSOFT CORP	26,169	0.1355	193,200
	2	ALPHABET INC	21,385	0.1194	179,131
	3	META PLATFORMS INC	15,075	0.1781	84,644
Manufacturing	1	NVIDIA CORP	8,965	0.3007	29,818
	2	APPLE INC	7,900	0.0539	146,636
	3	INTEL CORP	5,430	0.0791	68,679
Mining	1	SLB LTD	302	0.0177	17,080
	2	BAKER HUGHES CO	133	0.0094	14,119
	3	CONOCOPHILLIPS	117	0.0099	11,861
Other Svcs	1	WW INTERNATIONAL INC	39	0.0109	3,590
	2	CINTAS CORP	23	0.0010	23,462
	3	FRONTDOOR INC	16	0.0109	1,472
Prof & Business Svcs	1	ACCENTURE PLC	4,571	0.0586	77,983
	2	BOOZ ALLEN HAMILTON HLDG CP	2,034	0.0481	42,244
	3	COGNIZANT TECH SOLUTIONS	1,732	0.0443	39,070
Real Estate	1	CBRE GROUP INC	229	0.0060	37,878
	2	EQUINIX INC	229	0.0412	5,565
	3	JONES LANG LASALLE INC	159	0.0046	34,665
Retail Trade	1	AMAZON.COM INC	14,668	0.0410	358,106
	2	WALMART INC	1,876	0.0075	250,219
	3	EBAY INC	818	0.0505	16,211
Transportation/Warehousing	1	UNITED PARCEL SERVICE INC	284	0.0032	89,810
	2	DELTA AIR LINES INC	280	0.0058	48,629
	3	UNITED AIRLINES HOLDINGS INC	269	0.0064	41,854
Utilities	1	NEXTERA ENERGY INC	191	0.0195	9,801
	2	NRG ENERGY INC	160	0.0125	12,814
	3	DUKE ENERGY CORP	132	0.0070	18,778
Wholesale	1	MCKESSON CORP	574	0.0266	21,541
	2	CARDINAL HEALTH INC	467	0.0214	21,859
	3	INSIGHT ENTERPRISES INC	318	0.0361	8,812

This table reports the three firms with the largest number of AI-related U.S. jobs within each sector (based on 2-digit NAICS industry codes) in 2024 for the U.S. public-firm sample. Rank is assigned separately within each industry. The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.19. Top 30 IT-Sector Firms by AI Job Share in 2024

Rank	Firm	AI share	Industry	AI jobs	Total jobs
1	INNODATA INC	0.7169	Prof & Business Svcs	595	830
2	C3.AI INC	0.5156	Information	960	1,862
3	COVEO SOLUTIONS INC	0.3553	Information	27	76
4	BAIDU INC	0.3434	Information	114	332
5	CERENCE INC	0.3067	Information	69	225
6	VERITONE INC	0.3037	Information	147	484
7	SOUNDHOUND AI INC	0.2965	Information	110	371
8	LIVEPERSON INC	0.2137	Information	178	833
9	REKOR SYSTEMS INC	0.2044	Information	37	181
10	UIPATH INC	0.1869	Information	386	2,065
11	META PLATFORMS INC	0.1781	Information	15,075	84,644
12	GENIUS SPORTS LTD	0.1745	Information	26	149
13	STRATEGY INC	0.1711	Information	253	1,479
14	MOBILEYE GLOBAL INC	0.1695	Information	20	118
15	EGAIN CORP	0.1548	Information	61	394
16	PONY AI INC	0.1542	Information	31	201
17	CYNGN INC	0.1429	Information	16	112
18	APPLOVIN CORP	0.1360	Information	144	1,059
19	MICROSOFT CORP	0.1355	Information	26,169	193,200
20	PINTEREST INC	0.1341	Information	810	6,042
21	APPLIED DIGITAL CORP	0.1329	Information	23	173
22	INUVO INC	0.1319	Information	19	144
23	SNOWFLAKE INC	0.1257	Information	954	7,591
24	FISCALNOTE HOLDINGS INC	0.1194	Information	64	536
25	ALPHABET INC	0.1194	Information	21,385	179,131
26	STAGWELL INC	0.1193	Prof & Business Svcs	109	914
27	CS DISCO INC	0.1178	Information	76	645
28	SNAP INC	0.1160	Information	746	6,431
29	COREWEAVE INC	0.1142	Information	87	762
30	BIGBEAR.AI HOLDINGS INC	0.1140	Information	75	658

This table reports the 30 IT-sector firms with the highest AI-related job shares in 2024 for the U.S. public-firm sample. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. The IT-sector sample includes firms in information (NAICS 51) and professional and business services (NAICS 54). The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.20. Top 30 IT-Sector Firms by Number of AI Jobs in 2024

Rank	Firm	AI jobs	AI share	Industry	Total jobs
1	MICROSOFT CORP	26,169	0.1355	Information	193,200
2	ALPHABET INC	21,385	0.1194	Information	179,131
3	META PLATFORMS INC	15,075	0.1781	Information	84,644
4	INTL BUSINESS MACHINES CORP	9,440	0.0900	Information	104,923
5	ACCENTURE PLC	4,571	0.0586	Prof & Business Svcs	77,983
6	SALESFORCE INC	4,433	0.0802	Information	55,245
7	ORACLE CORP	4,037	0.0453	Information	89,111
8	ADOBE INC	2,775	0.0996	Information	27,873
9	VERIZON COMMUNICATIONS INC	2,465	0.0253	Information	97,536
10	AT&T INC	2,259	0.0158	Information	142,681
11	BOOZ ALLEN HAMILTON HLDG CP	2,034	0.0481	Prof & Business Svcs	42,244
12	SERVICENOW INC	1,904	0.0999	Information	19,067
13	INTUIT INC	1,902	0.0789	Information	24,109
14	COGNIZANT TECH SOLUTIONS	1,732	0.0443	Prof & Business Svcs	39,070
15	COMCAST CORP	1,457	0.0179	Information	81,392
16	SAP SE	1,300	0.0543	Information	23,939
17	DEUTSCHE TELEKOM	1,243	0.0167	Information	74,584
18	WORKDAY INC	980	0.0577	Information	16,982
19	C3.AI INC	960	0.5156	Information	1,862
20	SNOWFLAKE INC	954	0.1257	Information	7,591
21	NETFLIX INC	931	0.0520	Information	17,897
22	PALO ALTO NETWORKS INC	875	0.0705	Information	12,416
23	INFOSYS LTD	856	0.0300	Prof & Business Svcs	28,504
24	PINTEREST INC	810	0.1341	Information	6,042
25	SNAP INC	746	0.1160	Information	6,431
26	THOMSON-REUTERS CORP	658	0.0478	Information	13,767
27	AUTODESK INC	643	0.0706	Information	9,111
28	ATLASSIAN CORP	613	0.0864	Information	7,091
29	GARTNER INC	612	0.0371	Prof & Business Svcs	16,507
30	INNODATA INC	595	0.7169	Prof & Business Svcs	830

This table reports the 30 IT-sector firms with the largest number of AI-related U.S. jobs in 2024 for the U.S. public-firm sample. AI-related jobs are U.S. jobs identified using keyword matches in job titles or job descriptions. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. The IT-sector sample includes firms in information (NAICS 51) and professional and business services (NAICS 54). The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.21. Top 30 Non-IT Firms by AI Job Share in 2024

Rank	Firm	AI share	Industry	AI jobs	Total jobs
1	XPENG INC	0.3645	Manufacturing	109	299
2	NVIDIA CORP	0.3007	Manufacturing	8,965	29,818
3	ABSCI CORP	0.2301	Manufacturing	55	239
4	COUPANG INC	0.2098	Retail Trade	214	1,020
5	NERDY INC	0.1835	Education Svcs	29	158
6	AMBARELLA INC	0.1833	Manufacturing	68	371
7	WERIDE INC	0.1765	Manufacturing	33	187
8	CEVA INC	0.1613	Real Estate	10	62
9	ZEPP HEALTH CORP	0.1594	Manufacturing	11	69
10	MERCEDES BENZ GROUP AG	0.1424	Manufacturing	198	1,390
11	PENGUIN SOLUTIONS INC	0.1362	Manufacturing	88	646
12	SES AI CORP	0.1321	Manufacturing	21	159
13	STEM INC	0.1308	Admin Support	62	474
14	ALIBABA GROUP HLDG	0.1197	Retail Trade	116	969
15	TAL EDUCATION GROUP	0.1194	Education Svcs	8	67
16	ONE STOP SYSTEMS INC	0.1149	Manufacturing	10	87
17	SUPER MICRO COMPUTER INC	0.1138	Manufacturing	279	2,452
18	NIO INC	0.1134	Manufacturing	39	344
19	TEMPUS AI INC	0.1134	Health Care	403	3,555
20	RECURSION PHARMA INC	0.1101	Manufacturing	97	881
21	ASTERA LABS INC	0.1082	Manufacturing	42	388
22	ADVANCED MICRO DEVICES	0.1081	Manufacturing	1,670	15,444
23	DIDI GLOBAL INC	0.1079	Transportation/Warehousing	26	241
24	ALLURION TECHNOLOGIES INC	0.1064	Manufacturing	15	141
25	SERVE ROBOTICS INC	0.1061	Transportation/Warehousing	21	198
26	KNIGHTSCOPE INC	0.1061	Admin Support	14	132
27	UPSTART HOLDINGS INC	0.1025	Finance/Insurance	199	1,941
28	CERIBELL INC	0.1000	Manufacturing	42	420
29	ALTO NEUROSCIENCE INC	0.0877	Manufacturing	10	114
30	BUTTERFLY NETWORK INC	0.0816	Manufacturing	32	392

This table reports the 30 non-IT firms with the highest AI-related job shares in 2024 for the U.S. public-firm sample. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. AI-related jobs are U.S. jobs identified using keyword matches in job titles or job descriptions. The non-IT sample excludes firms in information (NAICS 51) and professional and business services (NAICS 54). The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.22. Top 30 Non-IT Firms by Number of AI Jobs in 2024

Rank	Firm	AI jobs	AI share	Industry	Total jobs
1	AMAZON.COM INC	14,668	0.0410	Retail Trade	358,106
2	NVIDIA CORP	8,965	0.3007	Manufacturing	29,818
3	APPLE INC	7,900	0.0539	Manufacturing	146,636
4	INTEL CORP	5,430	0.0791	Manufacturing	68,679
5	JPMORGAN CHASE & CO	4,527	0.0236	Finance/Insurance	191,930
6	CISCO SYSTEMS INC	3,972	0.0582	Manufacturing	68,223
7	CAPITAL ONE FINANCIAL CORP	3,284	0.0458	Finance/Insurance	71,757
8	DELL TECHNOLOGIES INC	3,163	0.0521	Manufacturing	60,732
9	WELLS FARGO & CO	2,699	0.0130	Finance/Insurance	206,854
10	TESLA INC	2,545	0.0424	Manufacturing	60,006
11	CVS HEALTH CORP	2,173	0.0165	Health Care	131,826
12	CITIGROUP INC	1,877	0.0207	Finance/Insurance	90,850
13	WALMART INC	1,876	0.0075	Retail Trade	250,219
14	BANK OF AMERICA CORP	1,790	0.0096	Finance/Insurance	185,569
15	ADVANCED MICRO DEVICES	1,670	0.1081	Manufacturing	15,444
16	QUALCOMM INC	1,546	0.0777	Manufacturing	19,907
17	UNITEDHEALTH GROUP INC	1,392	0.0189	Finance/Insurance	73,651
18	HP INC	1,333	0.0652	Manufacturing	20,436
19	LOCKHEED MARTIN CORP	1,323	0.0133	Manufacturing	99,685
20	VISA INC	1,316	0.0677	Finance/Insurance	19,450
21	GENERAL MOTORS CO	1,269	0.0178	Manufacturing	71,209
22	JOHNSON & JOHNSON	1,251	0.0268	Manufacturing	46,766
23	ELEVANCE HEALTH INC	1,177	0.0223	Finance/Insurance	52,772
24	AMERICAN EXPRESS CO	1,138	0.0297	Finance/Insurance	38,326
25	FORD MOTOR CO	1,082	0.0217	Manufacturing	49,831
26	CIGNA GROUP (THE)	1,065	0.0228	Finance/Insurance	46,702
27	MORGAN STANLEY	1,053	0.0151	Finance/Insurance	69,844
28	PFIZER INC	1,043	0.0214	Manufacturing	48,678
29	PAYPAL HOLDINGS INC	1,018	0.0647	Finance/Insurance	15,746
30	BROADCOM INC	992	0.0322	Manufacturing	30,813

This table reports the 30 non-IT firms with the largest number of AI-related U.S. jobs in 2024 for the U.S. public-firm sample. AI-related jobs are U.S. jobs identified using keyword matches in job titles or job descriptions. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm. The non-IT sample excludes firms in information (NAICS 51) and professional and business services (NAICS 54). The 2024 firm-year sample is restricted to firms with at least 50 U.S. workers observed in Revelio, at least 50 employees in Compustat, and to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.23. Correlations between Cognism and Revelio AI Measures

Year	Number of Firms	Corr. AI share	Corr. AI job count	Corr. total job count
2010	668	0.450	0.771	0.873
2011	710	0.517	0.837	0.876
2012	753	0.357	0.811	0.886
2013	832	0.351	0.839	0.888
2014	832	0.427	0.871	0.904
2015	856	0.511	0.873	0.910
2016	877	0.629	0.854	0.912
2017	929	0.515	0.918	0.920
2018	940	0.717	0.929	0.926
2019	922	0.477	0.936	0.927
2020	848	0.784	0.931	0.932
2021	598	0.723	0.929	0.937

This table reports year-by-year correlations between two resume data sources, Revelio and Cognism, across three different variables, covering 2010 to 2021. Each row corresponds to a calendar year. The "Number of Firms" column gives the number of underlying firms on which each year's correlations are computed. The remaining three columns report Pearson correlation coefficients between the Revelio and Cognism measures of, respectively: the AI share (the share of jobs classified as AI-related), the AI count (the absolute number of AI-related jobs), and the total job count (the overall number of jobs).

Table A.24. AI Job Functional Taxonomy

Category	Definition
AI Research & Model Development	Builds or directly advances AI models or core algorithms, including machine learning, deep learning, NLP, computer vision, reinforcement learning, and applied AI research.
AI Data, Annotation & Evaluation Support	Creates, curates, labels, validates, or analyzes data and evaluation assets primarily to support AI model training, testing, or improvement.
AI Platform, Infrastructure & MLOps	Builds or operates the technical platforms, pipelines, infrastructure, and deployment systems used to train, serve, monitor, or scale AI workloads.
AI Application & Systems Engineering	Integrates AI models or methods into software, devices, products, or operational systems where the core job is engineering AI-enabled functionality rather than creating the underlying model.
AI Product, Program & Delivery Management	Leads product strategy, program execution, or cross-functional delivery for AI products, AI features, or AI implementation without being the primary hands-on model builder.
AI Governance, Risk & Responsible AI	Sets, enforces, audits, or advises on AI-related governance, compliance, ethics, privacy, security, model risk, or responsible AI practices.
AI Business Leadership & Functional Oversight	Provides executive, business-unit, or functional leadership over AI-related teams, portfolios, or investments where the person oversees AI direction but is not primarily doing hands-on AI work.
Domain Analytics & Predictive Modeling	Uses statistical analysis, forecasting, optimization, or machine learning mainly to generate business, scientific, clinical, financial, or operational insights rather than to build AI products or platforms.
AI Go-to-Market, Solutions & Advisory	Sells, solutions, markets, evangelizes, or advises on AI-related products or transformations for customers or internal stakeholders without being the primary builder.
General Digital/Technical Role with AI Tooling or Claims	Works in a broader software, operations, consulting, or business role where AI is presented as one tool, initiative, or supporting claim but not the central function of the job.
Incidental or Unsubstantiated AI Mention	Mentions AI-related terms but provides insufficient evidence that AI work, AI deployment, or meaningful AI use was a substantive part of the role.

This table defines the functional categories used to describe what employees do within the AI-related job sample. The categories group sampled AI-related jobs by the worker's primary role in the AI ecosystem. The taxonomy is constructed from a random sample of 4,500 AI-related jobs, with 300 jobs per year from 2010 through 2024. The categories are mutually exclusive and are based on the worker's job title and self-reported job description rather than industry, employer, or seniority. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.25. Distribution of AI Jobs Across Functional Categories

Category	N	% jobs
AI Research & Model Development	650	14.4
AI Data, Annotation & Evaluation Support	51	1.1
AI Platform, Infrastructure & MLOps	273	6.1
AI Application & Systems Engineering	469	10.4
AI Product, Program & Delivery Management	529	11.8
AI Governance, Risk & Responsible AI	61	1.4
AI Business Leadership & Functional Oversight	491	10.9
Domain Analytics & Predictive Modeling	531	11.8
AI Go-to-Market, Solutions & Advisory	334	7.4
General Digital/Technical Role with AI Tooling or Claims	775	17.2
Incidental or Unsubstantiated AI Mention	336	7.5

This table reports the distribution of the 4,500 sampled AI-related jobs across the taxonomy of functional AI job categories. Each job is assigned to exactly one category using the worker's job title and job description. "N" reports the number of jobs assigned to each category, and "% jobs" reports the corresponding share of the classified sample. Categories are mutually exclusive. Table A.24 provides the definitions for functional categories. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table A.26. Share of AI Jobs by Aggregated AI Functional Category

	All	Tech	Non-tech
Core AI development	45.2%	42.9%	47.9%
Management	22.7%	23.1%	22.2%
Support for AI products/services	7.4%	11.6%	2.5%
Other	24.7%	22.4%	27.4%
Jobs	4,500	2,420	2,080

This table reports the distribution of 4,500 sampled AI-related jobs across aggregated taxonomy categories. Each job is assigned to exactly one category using the worker's job title and self-reported job description, which we then aggregate into four groups. "Core AI development" combines "AI Research & Model Development", "AI Data, Annotation & Evaluation Support", "AI Platform, Infrastructure & MLOps", "AI Application & Systems Engineering", "AI Governance, Risk & Responsible AI", and "Domain Analytics & Predictive Modeling". "Management" combines "AI Product, Program & Delivery Management" and "AI Business Leadership & Functional Oversight". "Support for AI products/services" consists of "AI Go-to-Market, Solutions & Advisory". "Other" combines "General Digital/Technical Role with AI Tooling or Claims" and "Incidental or Unsubstantiated AI Mention". Table A.24 provides the definitions for functional categories. Column 1 includes all sampled jobs, and columns 2 and 3 split firms into the tech sector—NAICS 51, 54, and semiconductor manufacturing (NAICS 334413)—and all remaining (non-tech) sectors. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

B. Organization Capital: Definition, Measurement, and Case Studies

B.1. Definition and Neighboring Concepts

Organization Capital (OC) is a durable, organization-specific intangible capability that improves how a firm combines inputs (people, tasks, routines, technologies, management systems, and physical assets), coordinates and governs activity, develops products, and brings output to market. It accumulates through learning (from experience, training, and management) and may be embodied in the organization itself or in key employees. As firms learn how to do things better, they slowly accumulate OC and become more productive. Once built, the asset is non-rival within the firm: a documented procedure or a structured record serves many employees, units, and tasks at once without depletion, so OC scales internally at little additional cost. However, it travels poorly outside the firm: routines and systems tuned to one firm lose much of their value elsewhere, which is what makes the asset hard to separate, transfer, imitate quickly, or value from accounting data.

These five interrelated diagnostics differentiate OC from other concepts:

- **Built (vs purchased).** OC is built up gradually through firms' learning-by-doing rather than bought on the market.
- **Durable (vs ephemeral).** OC represents repeatable useful knowledge rather than one-off outputs; it affects current and future production.
- **Organization-specific (vs general).** Its usefulness is highest in the firm in which it was accumulated: OC is less *portable* and *marketable* than some other intangible assets, such as patents. Organization specificity also tends to make OC relatively excludable.
- **Embedded (vs separable).** It resides in platforms, systems, routines, communication channels, culture, processes, workflows, and—to the extent it has not been written down—in workers themselves. This characteristic also differentiates OC from inputs into intangible capital building (e.g., R&D expense) and its outputs (e.g., patents).
- **Valuable (vs not).** Its use raises the productivity of inputs and market-facing activities by improving innovation, coordination, execution, production, and service. Knowledge that ceases to be useful—because the environment it describes changed—depreciates out of the stock of OC.

The existing literature often lists the following recurring families of OC examples, although they are often vague due to the difficulty of identifying and measuring OC in practice:

- **Information, learning, and matching:** personnel information, worker–task and worker–worker matches, task information, built-up plant expertise, and knowledge of how to use technology.
- **Human capital and key talent:** firm-specific skills, on-the-job training, informal learning, managers, engineers, product managers, software engineers, and other essential personnel.

- **Management and governance:** management practices, control systems, culture, relational contracts, leadership, capabilities, and management structures.
- **Processes and systems:** routines, internal communication, distribution, supply chains, systems (e.g., Enterprise Resource Planning, HR, Customer Relationship Management), platforms (e.g., cloud, enterprise application, developer and software platforms), knowledge sharing, and operational excellence.

In terms of how to conceptually separate OC from related intangible investments and capital concepts, we classify the durable OC stock separately from *inputs* (e.g., SG&A expenditures) or *efforts* to change something that creates OC and from *outputs* (product, patent, customer relationship) that OC enables. R&D spending, IT purchases, sales and marketing, training, process redesign, and co-invention are expenditures, activities, and inputs. They may create several different outputs. OC is the durable, firm-specific capability that persists after those activities are complete.

R&D and knowledge capital. R&D may denote an expenditure flow, a portfolio of projects, or the scientific and technical knowledge produced by those projects. In standard intangible-capital taxonomies, patents, formulas, designs, algorithms, experimental findings, and technical know-how are knowledge capital, whereas OC is the firm-specific system that selects, combines, coordinates, and deploys that knowledge (Corrado et al., 2009; Peters and Taylor, 2017; Eisfeldt et al., 2026a; Charoenwong et al., 2026). The two stocks are nevertheless complementary: both are cumulative, path-dependent, hard to observe, and often embodied in skilled labor. R&D can also build an organizational learning capability thus creating OC (Cohen and Levinthal, 1989). The practical boundary is the locus and portability of value. A patent, molecule, design, or licensable code base is normally knowledge capital that is portable across firms; research-portfolio governance, cross-functional product-development routines, scientific communication networks, and architectural competence are OC when their value depends on the firm's configuration. AI makes this distinction especially sharp: Cockburn et al. (2019) argue that deep learning is a general-purpose method of invention, and its value depends less on the largely public algorithms themselves than on the proprietary data and firm-specific research processes built around them—precisely the kinds of assets that constitute OC. Note also that OC need not be novel, whereas knowledge capital typically is. R&D and OC are therefore adjacent and complementary, not synonymous.

Information technology. The same distinction applies to IT. Hardware, software, databases, networks, and cloud infrastructure are technical or digital capital; OC is the complementary organizational asset that determines how effectively they are used. Productivity gains arise when technology is combined with decentralization, skilled labor, training, redesigned work practices, and complementary organizational investment (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002; Brynjolfsson et al., 2002). A packaged enterprise system is not OC merely because it is intangible: it is bought rather than accumulated, and the same system runs at any firm that licenses it, so it fails both the accrued and the organization-specific diagnostics. Custom code and proprietary data are the harder case, and the storage-medium criterion resolves it: a licensable code base is portable digital capital,

whereas a data asset assembled from the firm's own operations, or an architecture that encodes its particular workflows, has little value elsewhere and is OC. Unambiguously OC, finally, are the authority structure, role definitions, exception-handling routines, training, and cross-unit coordination that make any of these systems productive. The fact that high-OC firms spend more on IT therefore indicates complementarity rather than identity (Eisfeldt and Papanikolaou, 2013). Even when the two assets are co-specialized, empirical work should measure them separately.

Absorptive capacity. Where IT is an adjacent technical asset, absorptive capacity is closer to an internal organizational capability. Cohen and Levinthal (1990) define it as the ability to recognize the value of external information, assimilate it, and apply it commercially. It is path-dependent, rests on prior related knowledge, and can be built through R&D, production experience, technical training, and organizational interfaces. In pharmaceuticals, effective absorption also depends on incentives to publish and on connections to the wider scientific community, not only on the volume of in-house research (Cockburn and Henderson, 1998). These features make absorptive capacity a durable, firm-specific capability embodied in scientists, communication channels, incentives, shared language, and the architecture linking internal and external knowledge. Consistent with this, Woeppel and Yavuz (2026) construct a firm-level measure of absorption capability from patent data and show that it predicts stronger innovation and growth. Its scope is nonetheless narrower than OC: OC also includes worker matching, production routines, culture, supply-chain coordination, governance, and other capabilities unrelated to external learning. Absorptive capacity is therefore best treated as a subtype or dimension of OC.

Co-invention. This stock-versus-process distinction is sharper for co-invention. Bresnahan and Greenstein (1996) use the term for the user-side experimentation, discovery, and complementary redesign needed to make a general technology economically valuable. Adopters alter applications, workflows, skills, and organizational arrangements, so co-invention often builds routines, process knowledge, communication patterns, and governance that are slow to accumulate and hard to imitate. Yet co-invention is not itself a capital category. A project may yield custom software or data, a redesigned service, a customer interface, and persistent coordination routines; of these, the persistent coordination routines are unambiguously OC, and the data may be as well, depending on whether it was assembled from and used in the firm's own operations. Co-invention is therefore to OC as investment is to capital: an important formation mechanism, not the stock itself.

Process and product innovation. The same logic organizes process and product innovation. A business process innovation is a new or significantly improved process that has been brought into use, such as a production method, logistics system, administrative process, or information flow (OECD and Eurostat, 2018). Process innovation is therefore an implemented event, while OC is a durable capability. The overlap is substantial because new work design, cross-functional production, internal communication, supply-chain management, and process improvement can become OC once they are learned, stabilized, and embedded in routines, roles, and tacit coordination. Process innovations are also often difficult to sell separately and generate returns tied to the firm's own scale (Cohen and Klepper, 1996). Still, a process change may reside mainly in new equipment or in a

purchased software package. The converse also holds: OC need not be innovative. An established routine that no competitor would find novel can be highly valuable to the firm that operates it, and OC can arise from the codification of practice a firm already followed. Novelty and durable firm-specific capability are separate properties; OC requires the second, not the first. OC is thus the persistent organization-specific stock that some (but not all) process innovations create—and that also accumulates without any innovation event at all.

A product innovation is a new or significantly improved good or service introduced to the market (OECD and Eurostat, 2018). Its immediate object is the offering: the underlying design, patent, formula, or software is normally knowledge or digital capital, and market introduction is an innovation event. OC nevertheless enables product innovation through repeatable product-development systems, cross-functional coordination, launch routines, quality systems, platform governance, and the ability to scale reliable delivery. The clean boundary separates the artifact from the capability: the product or patent is not normally OC, but the system for repeatedly generating, integrating, launching, and scaling products can be.

Thus, OC can support both process and product innovation even though its own locus is closer to the persistent organizational capability than to the innovation event itself.

Architectural competence and architectural innovation. Finally, architectural competence and architectural innovation show why similarly named concepts can occupy different categories. Henderson and Cockburn (1994) distinguish local component competence from architectural competence in pharmaceutical research. Architectural competence integrates disciplinary knowledge, coordinates across programs, filters information, and develops new competencies; it is embodied in communication channels, managerial systems, incentives, norms, and the organization of research. It is collective, firm-specific, durable, and productive through integration rather than through a separable piece of technical knowledge. The best classification is therefore an R&D-specific subtype of OC, with a narrower domain than the firm-wide concept.

Architectural innovation is a different object. It is a product-system innovation that changes linkages among components while leaving core component concepts largely intact (Henderson and Clark, 1990). Established firms can fail because their communication channels and information filters embody the old architecture. The innovation event is not OC; rather, it reveals the value and rigidity of the architectural knowledge already embedded in the organization. It devalues existing OC and requires the accumulation of new OC.

Customer capital. A different boundary separates internal supply-side capabilities from demand-side assets. Customer capital is a durable intangible asset built through sales and marketing, customer service, brand-building, customer information, acquisition, retention, and related strategies (He et al., 2024). Like OC, it is firm-specific, built gradually through people and routines, frequently expensed rather than capitalized, important for scale, and non-rival in the sense that it can serve many transactions at once. Broad definitions of OC may therefore include customer-facing capabilities that are tightly integrated with the business model. Under a narrower taxonomy, however, OC concerns how the firm

coordinates, innovates, and produces, whereas customer capital concerns who buys, why customers return, and the value of relationships, reputation, and brand. Customer-service routines may be OC, while the relationships and reputation those systems help build are customer capital. This internal-versus-external distinction is also central to efforts to separate OC from brand and customer assets within SG&A (Chawla, 2025; Eisfeldt et al., 2026a; Charoenwong et al., 2026).

Summary classification. Table B.1 applies the same rule across all neighboring concepts. The decisive questions are: What is the durable stock? Where does its value reside? How portable is it outside the firm? Expenditures, redesign activities, and innovation events should be classified separately from the persistent capabilities they create.

Table B.1. Organization capital and neighboring concepts

Concept	Relation to OC	Key boundary
R&D / knowledge capital	*Adjacent stock*	Scientific and technical knowledge is knowledge capital. The organization of R&D—interfaces, incentives, cross-functional routines, and architectural competence—is OC when its value depends on the firm's configuration.
IT / digital capital	*Adjacent stock*	Hardware, packaged software, and networks are technical assets, as is portable code. Data and architecture built from the firm's own operations, and the work practices, authority, training, and process redesign that make IT productive, are OC.
Absorptive capacity	*OC subtype*	A path-dependent capability for identifying, assimilating, and using external knowledge. It fits within OC but is narrower than the firm-wide concept.
Co-invention	*Formation process*	User-side experimentation and redesign may create OC, digital capital, customer capital, or product outputs. Co-invention is a formation activity, not the resulting stock.
Process innovation	*Event → possible OC*	A new process is an innovation event. When it becomes a persistent, firm-specific routine or coordination system, the resulting stock is OC.
Product innovation	*Output enabled by OC*	The product, patent, or design is not normally OC. Repeatable product-development, launch, quality, and scaling capabilities can be OC.
Architectural competence	*OC subtype*	Integrative research competence is a close, domain-specific example of OC.
Architectural innovation	*Event*	Reconfiguring product linkages is an innovation event. It can devalue old OC and require new OC, but it is not the stock itself.
Customer capital	*Adjacent stock*	Customer relationships, brand, service, and retention assets are demand-side capital. The internal systems used to build them may be OC.

B.2. Measuring OC Using Job Descriptions

We propose a new measure of OC using workers' job descriptions based on whether an individual job builds OC. The key advantage of this novel job-description measure is

that it observes worker activities that build OC at the worker-job level—for example, codifying firm-specific knowledge into reusable systems, routines, and processes—rather than inferring OC from aggregate expenditures. This makes it substantially more informative than capitalized-SG&A measures, which capture an input into OC but cannot distinguish spending that actually builds durable organizational capabilities from unrelated SG&A spending. Relative to measures based on employee reviews of their employers—which capture employees' perceptions of the realized stock of culture, coordination, and management quality—the job-level measure identifies where and how OC is produced, allowing researchers to attribute its accumulation to particular workers, tasks, and technologies such as AI.

To develop our methodology for identifying OC-building jobs, each co-author independently built their own prompt and used different LLMs to classify jobs using the same random sample of 2,000 jobs. After comparing the outputs and the manual review of classifications, we converged on a prompt (copied below) that offered the highest accuracy and generalization in classification.

Our prompting approach has the following desirable features. First, it is agnostic on whether a job is AI-related or not: the prompt does not contain any mentions of AI or how AI jobs might be more or less likely to be building OC. This helps to *not* bias the results on classifying jobs into OC-building based on how related they are to AI. Second, our definition of OC in the prompt comes from the core seminal papers (Prescott and Visscher, 1980; Atkeson and Kehoe, 2005; Eisfeldt and Papanikolaou, 2013): this allows for a high-level conceptual but precise-enough definition of key features of OC outlined in Part B.1 above. We find this approach attractive since it does not rely on rough proxies (e.g., occupations, seniority, job titles, industry) for what types of jobs should build OC: a manager can oversee a few workers without building any OC or, instead, can lead and coordinate across teams to build systems to improve how firms operate—a good example of building OC. Instead, the model reads each job's description and title and identifies whether the job has the key features of building OC and offers substantial evidence of OC-building. Finally, we test our final prompt across several frontier LLMs and find highly consistent outputs across multiple runs, with approximately 95% agreement across models.

To classify jobs as OC-building or not, we first build two samples of jobs from the Revelio Labs resume data. The AI sample includes *all* 593,332 AI-related jobs in our public-firm sample with non-empty job descriptions (about 91% of all AI-related jobs). For cost-efficiency, the non-AI sample consists of a representative random sample of 15,000 non-AI jobs with non-empty job descriptions, drawn from the same firm-years as the AI jobs (1,000 jobs per year during 2010–2024). Table B.3 compares the 15,000-job random sample to the full population of non-AI jobs in terms of industry (Panel A) and firm size (Panel B) distribution. In Panel A, manufacturing accounts for 27.02% of all non-AI jobs (and 27.83% in the random sample), finance and insurance accounts for 17.67% of the population (and 16.14% of the sample), and retail trade for 14.64% (14.60%). Overall, across industries, the population and the random sample employment shares line up closely. In Panel B, the median firm has 23,457 employees in the population and 24,207 in the random sample, and the respective means are 52,280 and 53,728. In the non-AI job sample, the firm size distribution matches almost exactly between the random sample and the whole

population.

To reduce variability in the LLM's output and improve reproducibility, we set the temperature to zero. In both samples (AI and non-AI jobs), we score each job with *GPT-5.4* which reads each job title and description and returns an OC score between 0 and 1. We classify a job as building OC if its score is at least 0.6, where 0.6 is the boundary between mixed evidence (0.4–0.6) and substantial evidence (0.6–0.85) on building OC. The full prompt is as below:

```
DEFINITION - ORGANIZATION CAPITAL

Organization capital is a durable, intangible, and firm-specific asset.
Organization capital is knowledge specific to a firm that helps it convert
inputs (e.g., labor, physical capital, data) into output more efficiently.
Firms accumulate this knowledge through experience and learning-by-doing from
their operations, and as they do, this knowledge helps them to operate more
efficiently and become more productive. The faster the learning process, the
more quickly firms accumulate organization capital and realize productivity
gains. Organization capital is partly embodied in the knowledge of key
employees and built by them. It is distinct from physical capital and from
general, portable human capital (skills a worker takes to any employer). In
sum, organization capital is firm-specific knowledge that is embodied in the
organization itself (in its routines, practices, systems, processes,
employees, databases); it is acquired through learning-by-doing and partially
built by its employees.

STAGE 1 - SCORE INDIVIDUAL JOBS

You are an expert labor economist classifying LinkedIn resume entries. Each
input is ONE position a worker listed on their LinkedIn profile: title_raw is
the worker's raw, self-entered job title, and description is the worker's own
free-text account of what they did in that role (not an employer ad).

TASK

Decide whether the work contributes to building the employer's ORGANIZATION
CAPITAL and rate how strongly on a 0-1 scale, using the definition above.

HOW TO APPLY THE DEFINITION

1. Map the described activities onto the definition; a job scores high only
to the extent its work builds firm-specific organization capital
(accumulating operational/production knowledge through experience, codifying
it in routines/systems/databases, or developing key employees' firm-specific
knowledge).
```

2. General or portable skills, one-off individual output (a single analysis, model, design, or code change), or routine operation of existing assets without building firm-specific knowledge scores low.

3. Rely on the definition rather than your own broader intuitions.

SCORING RUBRIC (0-1)

0.00-0.15 Unrelated.

0.16-0.40 Incidental.

0.41-0.60 Mixed.

0.61-0.85 Substantial (building firm-specific organizational capability is central).

0.86-1.00 Core (role exists primarily to build firm-specific organization capital).

OUTPUT

Return ONLY JSON with org_score (number 0.00-1.00). Judge the WORK described, not the title alone. Ignore mojibake/formatting artifacts. Do not invent activities not supported by the text.

STAGE 2A - NAME MECHANISMS FOR JOBS THAT BUILD ORGANIZATION CAPITAL

Below are JOB DESCRIPTIONS grouped by clustering. They share a common mechanism for HOW the work BUILDS the employer's ORGANIZATION CAPITAL.

Using the definition above, write ONE sentence naming the shared ECONOMIC MECHANISM by which these jobs build firm-specific organization capital. Examples include codifying firm-specific knowledge into systems/routines, developing key employees' firm-specific knowledge, or building reusable firm data assets.

Describe the reason, NOT the occupation or industry.

Return ONLY JSON: {"mechanism":<one sentence>}.

STAGE 2B - NAME MECHANISMS FOR JOBS THAT DO NOT BUILD ORGANIZATION CAPITAL

Below are JOB DESCRIPTIONS grouped by clustering. They share a common reason WHY the work does NOT build the employer's ORGANIZATION CAPITAL.

```
Using the definition above, write ONE sentence naming the shared ECONOMIC
REASON these jobs do NOT build firm-specific organization capital. Examples
include general/portable skills, one-off individual output, or routine
operation of existing assets.

Describe the reason, NOT the occupation or industry.

Return ONLY JSON: {"mechanism":<one sentence>}.

STAGE 3 - STANDARDIZE MECHANISM TAXONOMY

You are standardizing a taxonomy of economic mechanisms for ORGANIZATION
CAPITAL.

Below are candidate mechanism statements, each summarizing a cluster of real
jobs with job counts. Merge them into between 20 and 40 DISTINCT standardized
categories. Each category is ONE full sentence stating the ECONOMIC REASON
the work [builds / does not build] firm-specific organization capital.

RULES

Merge statements that express the SAME economic reason even if they come from
different industries or occupations.

Do NOT base categories on occupation or industry.

Do NOT create a catch-all "other" or "miscellaneous" category.

If one reason would dominate the data, SPLIT it into distinct sub-mechanisms.

Ground every category in the definition of organization capital above.

Return ONLY JSON: {"categories":[{"category_id":"[prefix]1","label":<one
sentence>}, ...]}.
```

One potential concern is that our OC classification simply picks up workers who write higher-quality job descriptions, but we do not find that to be the case. To assess this, we construct some simple writing-quality measures. OC jobs are about twice as long, but they use a more repetitive vocabulary (78 distinct words among the first 100, versus 84), cite fewer specific numbers and tools (7.9 versus 9.3 mentions per 100 words), and contain spelling errors at essentially the same rate as other AI job descriptions (7.1 versus 6.5 errors per 100 words). When comparing descriptions of similar length, the two groups are essentially indistinguishable across all quality measures.

We find that AI jobs are more likely to build OC than non-AI jobs. Table B.4 shows that 58% of AI jobs are OC-building versus 31% of non-AI jobs, and AI jobs have a higher mean OC score of 0.60 versus 0.42 for non-AI jobs. This pattern holds across the whole score

distribution—the largest share of AI jobs falls in the "substantial" band, while the largest share of non-AI jobs falls in the "incidental" band.

We next look at how jobs that build OC differ from those that do not across several dimensions. First, we examine word clouds of job descriptions for OC-building and non-OC-building jobs, separately for non-AI and AI jobs, in Figure B.9. Among AI-related jobs, intuitively both types of jobs have *data* and *ai* as keywords, but OC-building positions emphasize *team*, *lead*, *product*, *business*, *design*, and *process*, combining language about leading teams and building business products and processes, whereas non-OC-building jobs are led by *model*, *learning*, and *develop*, emphasizing technical model development and deployment. Among non-AI-related jobs, OC-building positions are led by *team*, *process*, *lead*, *management*, and *business*, highlighting operational and managerial routines, whereas non-OC-building jobs are led by *customer*, *service*, *sale*, *provide*, *team*, *product*, *ensure*, emphasizing customer-facing service, sales, and support work. Comparing OC-building jobs across the AI and non-AI divide, both groups emphasize *team*, *lead*, and *business* as well as *product* and *process*, but AI-related positions also emphasize *data*, *ai*, and *model*, whereas non-AI-related positions stress *management* and *project*.

Second, we find that non-tech firms have significantly higher shares of AI jobs that build OC: 65.1% versus 50.6% in tech firms (defined as firms in NAICS 2-digit sector 51 or 54, or in semiconductor manufacturing).[61] These differences are consistent with the evidence in Appendix Table A.26 that, compared to tech firms, non-tech firms are more likely to employ AI-skilled workers in core AI development and less likely to have AI jobs selling and consulting on AI products and services: tech firms tend to deploy AI talent to build products for others, while non-tech firms use it to build capabilities in-house.

Third, consistent with prior work that argues that workers in management roles are likely to build OC (Eisfeldt and Papanikolaou, 2013; Dessein and Prat, 2022), we find that 79% of AI-related workers in management roles are classified into OC-building roles for all firms (Panel A of Table B.6), and this ratio is higher for non-tech firms (86%, Panel B) than for tech firms (73%, Panel C). Consistent with the idea that customer service roles are more likely to serve customers and are less likely to build durable organizational capabilities, only 22.6% of AI workers in "Support for AI products/services" build OC. Interestingly, among the group of workers who are more heavily involved in core AI development, about half are tagged as OC-building (50% for non-tech and 43% in tech firms), consistent with the discussion above that tech-facing roles do not automatically build OC.[62]

These patterns are sharper at the occupation level, where the standard occupation classifications provided by Revelio offer external validation on which occupations build OC. In Appendix Tables B.7 and B.8, we report the thirty occupations with the highest OC-building shares among AI jobs (computed as the number of OC-building AI jobs divided by the total number of AI jobs in each occupation) and the thirty occupations with the largest numbers of OC-building AI jobs in 2024. Management and business-

[61] These OC-building shares are computed from the full sample of 593,332 AI jobs with a job description over 2010–2024.

[62] The OC-building statistics in this table are based on the random sample of 4,500 AI jobs classified into AI job categories, and therefore slightly differ from those numbers above that are computed on the full sample of 593,332 AI jobs.

operations occupations make up 22 of the 30 occupations with the highest OC shares. Eight of the top ten are management occupations, including General and Operations Managers (94%), Training and Development Managers (92%), and Supply Chain Managers (92%). In contrast, the largest numbers of OC AI jobs are in technical occupations such as Software Developers and Data Scientists. Their OC shares, 56% and 57%, are close to the sample average. Thus, management occupations are the most intensively OC-building occupations, while technical occupations supply more OC-building AI jobs due to scale. Moreover, these statistics alleviate the concern that our LLM classification simply distinguishes genuine internal development roles from incidental mentions, sales and support roles, and sparse descriptions—that is, that it captures signal quality in the AI measure rather than organization capital per se.

We next examine the time-series dynamics. In Figure 6, we find that the majority of AI jobs build OC, and this share rises from about one-half in 2010 to around two-thirds in 2024 (panel (a)), both within the tech sector and outside of it (panel (b)). In contrast, only about one-third of non-AI jobs build OC (panel (c)), and this share is flat over the entire period in both subsectors (panel (d)). Appendix Figure B.10 shows that the same pattern holds using the average continuous OC score, confirming that the trend is not an artifact of the classification threshold. The stability of the time series of non-AI OC-building jobs indicates that the rise in the OC-building share of AI jobs reflects a genuine change in the nature of AI work rather than a more general trend of jobs becoming more OC-building or a drift in job descriptions of AI-related jobs. Consistent with the changing nature of AI work over time, we find that 57% of AI jobs in the earlier Wave I categories (ML, NLP, and CV) build OC, compared with 66% of jobs in the more recent Wave II categories (generative and agentic AI). More recent vintages of AI technologies appear increasingly oriented toward building OC, with the latest vintage—agentic AI—having the highest share of OC-building jobs (71%) among all AI vintages.

We then move beyond whether a job builds OC to *how* it does so: two jobs with identical OC scores may build very different kinds of firm-specific assets—one codifying tacit knowledge into systems and routines, another accumulating reusable data assets, a third building expertise in product innovation. Characterizing these mechanisms lets us ask not only whether AI work is more OC-intensive than other work, but whether it builds OC through different channels. We therefore assign each job a standardized reason category alongside its OC score. For OC-building jobs, the category records the mechanism through which the work builds firm-specific OC; for non-OC-building jobs, it records why the work does not—because the skills are general and portable, because the output is one-off and individual, or because the role operates existing assets rather than creating new ones. The LLM first induces a standardized set of mechanisms from the data and then assigns each job to its nearest category based on the job description. Appendix Table B.5 provides detailed definitions of each reason category.

Table 1 shows how AI and non-AI jobs are distributed across the reason categories within each group—OC-building and non-OC-building. Panel A shows that all OC-building job categories involve building systems, platforms, workflows, routines, and firm-specific knowledge and models. These categories also provide a sense of the types of assets created through OC-building jobs. Among these categories, AI jobs are concentrated in codifying data, models, and proprietary knowledge into reusable technical systems:

data platforms (9.1% of all AI jobs), proprietary knowledge generative-AI/NLP (6.9%), reusable AI/ML platforms (4.8%), and firm-specific predictive and decision models (2.3%). In contrast, non-AI-related jobs are more operational in nature, concentrating in hiring and talent development (5.8%) and manufacturing & process systems (3.5%). Both AI and non-AI jobs rank cross-functional operating routines among their top five categories (6.3% and 4.8%, respectively), highlighting the importance of within-firm coordination as a source of OC accumulation.

It is worth emphasizing the diversity of the types of OC built by both AI and non-AI workers. For example, AI workers build product and design systems (4.7%), marketing and revenue systems (4.1%), cloud, infrastructure and reliability operations (3.4%), financial and risk decision systems (3%), customer support systems (2%), security, compliance and risk systems (1.9%), R&D and scientific platforms (1.4%), and other operating processes. These activities represent the creation of durable organizational capabilities rather than one-off analytics, portable algorithms, or standardized routine work. The diversity of OC assets built by workers also underscores the difficulty of measuring OC using coarse signals like occupations, job seniority, or departments, since workers across each of these categories can contribute to its accumulation.

In Panel B, which covers jobs that do not build OC, AI jobs are concentrated in portable analytics and one-off models (12.6% of all AI jobs), broadly applicable research and algorithms (8.9%), and bespoke client consulting and implementation (7.8%). By contrast, non-AI jobs that do not build OC are overwhelmingly concentrated in standardized operations and support (37.9%), bespoke client consulting and implementation (8.8%), and transactional sales, marketing and account management (7.6%).

Appendix Table B.9 provides examples of AI-related job descriptions in our sample that do and do not build OC, together with the reason for each classification. OC-building AI jobs build systems the firm continues to use, such as data platforms, evaluation frameworks, and copilots built on internal workflows. Other AI jobs apply similar technical skills to one-time analyses, general research, client work, and products and sales. These examples illustrate that the LLM classifies an AI-related job as OC-building only when there is *sufficient evidence* that the work results in *durable* organizational capabilities—it does not assume that AI work is OC-building by default. For example, the following jobs are—correctly—not classified as OC-building. First, a "Research Scientist" is "responsible for developing advanced and reliable industrial *software* for automation applications such as smart manufacturing, industrial robots and cyber physical systems." Although the job involves building software, as we discuss above in Part B.1, software itself does not automatically qualify as OC. Second, a "Software Engineer" uses computer vision and "designs and implements a *method* of hiding and retrieving imperceptible patterns in images displayed on computer monitors and projection displays." Here, the method is an output of the research process and may or may not ultimately be implemented in the firm's systems or workflows.

Appendix Table B.10 provides analogous examples for non-AI-related jobs. OC-building non-AI jobs build the systems and practices through which the firm operates, including standardizing procedures, codifying safety and risk programs, training employees, and coordinating and collaborating across teams. In contrast, other non-AI jobs perform day-to-day tasks within existing systems or use skills that workers can readily

transfer to other employers. The examples also highlight differences in the types of OC built across the two samples. OC-building AI jobs primarily codify the firm's data, models, and proprietary knowledge into technical systems, whereas OC-building non-AI jobs more often build operational practices such as procedures, quality programs, and training, consistent with the distribution of reason categories in Table 1 (see Appendix Table B.5 for definitions).

B.3. Case Studies of AI Building Organization Capital

Organization capital (OC) is difficult to observe directly, which is why it is usually inferred from spending or from residuals; the case studies below take the opposite approach, using public accounts and detailed job descriptions of workers, to show what building OC actually looks like inside particular companies. The eight firms below span a wide range of industries—finance, consumer goods, biotech, home improvement retail, pharmacy, telecommunications, athletic apparel, and online travel—and an equally wide range of what AI is actually doing inside them. Morgan Stanley codifies advisor and underwriting judgment from research documents and client meetings; Procter & Gamble builds simulatable models of its plants from process telemetry; Moderna converts a short experimental history into predictive models while its organizational routines are still forming; Home Depot puts the expertise of store associates into a tool available in every aisle; CVS Health turns handwritten charts and claims into a searchable clinical record and a standing review routine; Verizon compresses the gap between its best and newest care representatives and field technicians; Nike turns costing and design judgment into tools and extends the record across a network of factories; and Expedia codifies how its own engineers diagnose incidents and ship software into agentic systems. The knowledge being converted is variously tacit clinical, commercial, mechanical, and engineering judgment, and the systems built to hold it range from digital twins and dispatch platforms to retrieval systems and standard operating procedures.

What these cases share is more striking than what separates them. In every instance the raw material is the firm's own operational record—the documents, transcripts, telemetry, claims, tickets, and transactions that accumulate as a by-product of doing business—rather than anything purchased or generally available; the same models, trained on public data, would build nothing firm-specific. In every instance the conversion is performed by the firm's own employees: the internal AI roles our measure captures are the labor that turns that record into a durable asset, and their descriptions show the work being done in the categories our classification identifies. And in most instances the resulting asset is non-rival within the firm and of limited use outside it—once the assistant, the twin, or the knowledge base exists, the cost of the next employee using it is close to zero, while much of its value would not transfer to a competitor. The industries and applications differ, but the underlying mechanism does not. Table B.2 summarizes which OC categories each case illustrates, and the coverage is wide: the eight firms together span seventeen of the eighteen distinct categories. Reusable data platforms and pipelines is the only category present at all eight—every one of them, whatever its industry, is converting its own operational record into a shared and reusable asset—and cross-functional operating routines is next at six.

Table B.2. Organization Capital Categories Represented in Each Case Study

	Category	Morgan Stanley	P&G	Moderna	Home Depot	CVS Health	Verizon	Nike	Expedia
O1	Reusable data platforms & pipelines	•	•	•	•	•	•	•	•
O2	Reusable AI/ML platforms & MLOps							•	•
O3	Firm-specific predictive & decision models				•				•
O4	Proprietary-knowledge GenAI/NLP systems	•	•	•		•			
O5	Reusable software platforms & workflows	•			•	•			
O6	Cross-functional operating routines	•		•	•	•	•	•	
O7	Reusable cloud, infrastructure & reliability operations						•		•
O8	Security, compliance & risk systems					•			
O9	Marketing & revenue-operations systems		•		•				•
O10	Customer-service & support systems						•		•
O11	Product & design systems							•	
O12	Manufacturing & process systems		•	•					
O13	Supply-chain & logistics systems		•					•	
O14	R&D & scientific platforms			•					
O15	Financial & risk decision systems	•							
O16	Hiring, training & talent development			•					
O18	Knowledge organization & ontologies						•		

Notes: A filled circle indicates that the category is illustrated in that firm's case study. The categories shown are, with few exceptions, among the five largest for the firm. Category codes and definitions follow Table 1 and Appendix Table B.5. Category O17 (Alliance & partnership systems) is omitted because this category is not represented among the firms in our case-study sample.

B.3.1. Morgan Stanley

Morgan Stanley offers an unusually well-documented illustration of how AI investment builds OC. The premise is that AI lets a firm convert the operational knowledge generated by its daily work—what its people know how to do, recorded in documents, transcripts, code, and client interactions—into a codified stock the firm owns and can redeploy. Morgan Stanley's head of firmwide AI, Jeff McMillan, has described five basic enterprise use cases for generative AI: the software development lifecycle, document management, translation, document creation, and marketing. Four of the five are exactly this conversion. Each takes a record of work the firm was already producing—code and design discussions, documents in digital and paper form, presentations and client materials, marketing content—and turns it into a reusable asset.

The advisor tools make the conversion concrete. The firm's AI Assistant, built on a corpus of roughly 100,000 proprietary research documents, reached over 98% adoption among financial-advisor teams, with document retrieval efficiency rising substantially (Morgan Stanley Wealth Management, 2024). The Debrief tool extends the same logic to a record that previously left no trace at all: with client consent it takes notes during meetings, summarizes key points, and writes a note into the firm's CRM. Advisor knowledge that had lived only in individual memory and ad hoc notes becomes a firm asset—the transformation Cullen et al. (2026) describe. It is also why the firm's own framing of its competitive position has shifted toward proprietary data and business logic as the foundation of what it sells (Son, 2026).

The AI jobs we observe at Morgan Stanley are the labor that performs this conversion into durable organizational assets. Our organization capital classification groups AI jobs not by application, as McMillan does, but by the kind of asset the work produces, so the conversion-oriented use cases above reappear below distributed across those categories. The descriptions cluster into the categories listed in Table 1 (and described in detail in Appendix Table B.5) that our classification identifies as most common at the firm:

- *Proprietary-knowledge GenAI/NLP systems—codifying documents and judgment rules into retrieval and generative systems.* The most visible body of work codifies the firm's documents and judgment rules into retrieval and generative systems. A GenAI engineer describes delivering "an enterprise GenAI Research & Policy Copilot (RAG + agent workflows) used by 600+ analysts/associates, handling 12K+ queries/week, reducing document search time by 35-45% across Wealth/Risk/Compliance teams." An executive director leads "the firmwide modernization of enterprise search," integrating "key platforms (e.g., HR, Social, Confluence) into Fusion and GenAI to deliver a unified knowledge experience." A senior business analyst serves as product owner for "AI Agent Assist call note summarization"—the Debrief capability—and an analyst describes "conversationalizing and orchestrating tools used by ~16,000 Financial Advisors into the AI @ Morgan Stanley Assistant chatbot."

- *Reusable software platforms & workflows—standardizing execution and reducing reliance on any individual employee.* The largest category turns tacit process and exception-handling knowledge into platforms other teams build on. A vice president describes an "AIMLOps Platform … redefining the firm's approach to AI/ML operationalization, including integration of Generative AI (GAI) and Large Language Models,"

with features for "agility, automation, scalability, reproducibility, CI/CD, monitoring," reporting a reduction in time-to-market of more than 35%. Another describes an internal GenAI platform serving "80,000+ Morgan Stanley employees." A head of client service change describes "delivering a knowledge management platform leveraging AI and AEM, re-architecting service analytics capabilities and implementing a CRM Agent Desktop solution using Salesforce."

- *Reusable data platforms & pipelines—unifying the records the rest depends on.* Codified data definitions and integration logic are the substrate the rest depends on. A senior data engineer describes designing "a cloud-native AI-powered investment management and regulatory data platform" that "unified trading, market, and risk data from Bloomberg, Reuters, ESG feeds, and OLTP systems." The Research & Policy Copilot above rests on an "end-to-end RAG ingestion pipeline" processing "2.5M+ documents (PDFs, policies, research notes)," which raised top-three retrieval precision from 62% to 82%.
- *Financial & risk decision systems—turning underwriting and control knowledge into standing decision routines.* Underwriting, fraud, and control knowledge is embedded into standing decision routines. A vice president in mortgage operations describes architecting "a GenAI-powered agent to analyze call transcripts and flag unaddressed items for mortgage Operations" and deploying "an LLM-based mortgage underwriting assistant to standardize underwriting memos, projected to save over 10,000 hours annually"—judgment that had resided only with individual staff becoming a firm-level standard.
- *Cross-functional operating routines—institutionalizing how complex work gets coordinated.* A smaller but persistent category. A vice president in product and technical program management describes leading "finance technology transformation, defining multi-year product vision, roadmap, and OKRs to modernize accounting platforms and streamline Wealth Management operations," delivering "an enterprise sub-ledger that reduced error rates by 40%."

Figure B.1 places these examples in the context of the firm's hiring. Panel (a) shows that both OC-building AI roles and other AI roles have grown steadily as a share of Morgan Stanley's US positions, with the growth accelerating after 2021; throughout the period the OC-building series runs roughly two to three times the size of the other, so the firm's AI hiring has been tilted toward work that converts proprietary operations into durable assets rather than toward generic technical capability, and that tilt has widened rather than narrowed as the level has risen. Panel (b) breaks that series into the reason categories in Table 1, showing the firm's five largest categories and grouping all remaining categories into "other." Reusable software platforms and workflows, together with reusable data platforms and pipelines, are present throughout and account for much of the level—they are the infrastructure on which anything else is codified—while financial and risk decision systems form a steady third component, consistent with the underwriting and compliance examples above. The most striking feature is the category of systems built on proprietary knowledge using generative AI and natural language processing, essentially

absent before the early 2020s and a substantial share by the end of the sample. That timing marks a change in what could be codified at all: retrieval over the firm's research library, summarization of advisor meetings, and standardization of written judgment were not feasible at scale under the earlier technology.

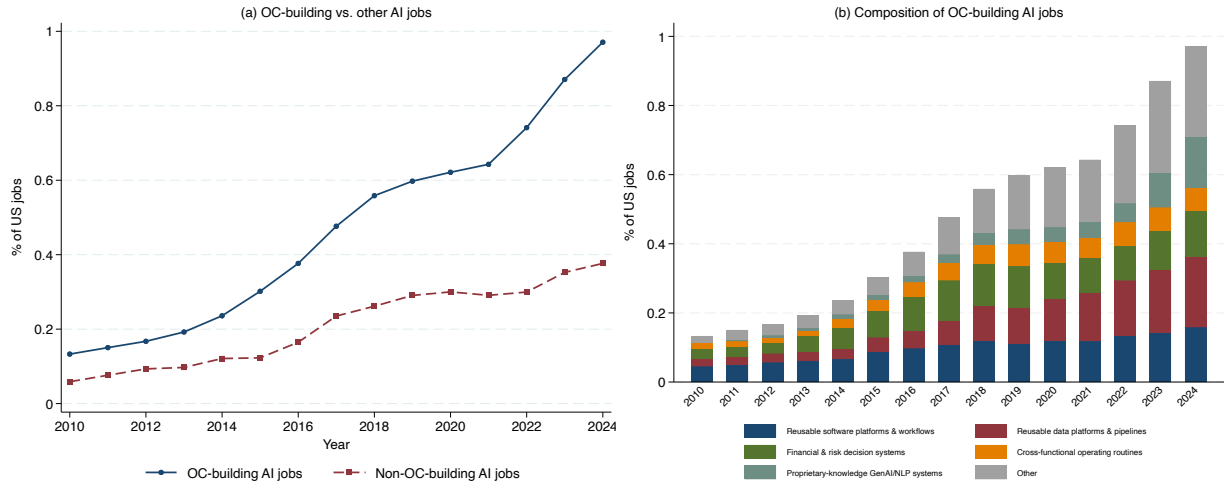


Figure B.1. AI jobs at Morgan Stanley as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification.

These uses of AI raise measured productivity through two channels. The first is that the same inputs yield more output: the research library, the advisors, and the client relationships are unchanged, but a far larger share of what the firm already knows is actually reachable at the moment it is needed. Codification also raises the floor rather than the ceiling, since a standardized underwriting memo makes the median underwriter perform closer to the best one, which is where the projected saving of over 10,000 hours comes from. The second channel is that the resulting asset is non-rival within the firm: once the Assistant exists, the cost of the sixteen-thousandth advisor using it is essentially zero.

Two features of the case are worth emphasizing. First, the codified stock is built out of operations rather than purchased: the assets being created are indexes of the firm's own research library, models trained on or fine-tuned to its own call transcripts, and workflow rules extracted from its own underwriting and origination processes. A large codified stock is not available for purchase as such; it accumulates through operating and recording. Second, not every AI use case builds OC. Translation is one of McMillan's five AI use cases but is generic—it converts nothing firm-specific into a durable asset—and consistent with this, a job in our data whose description centers on portable machine-learning methods applied to project-specific outputs, including machine translation, is classified as not OC-building. The distinction between AI that raises a firm's capability in general and AI that codifies what a firm already does is one our classification makes—as intended.

B.3.2. Procter & Gamble

Procter & Gamble illustrates the same logic in a very different setting. A consumer goods manufacturer's accumulated advantage lies in how it runs plants, moves product, and learns what consumers want—knowledge that has historically lived in the experience of process engineers, planners, and brand managers. AI lets the firm convert more of that operational knowledge into a codified stock it can query, simulate against, and redeploy across sites.

The clearest instance is the digital twin. Rather than halting a production line to trial a process change, P&G runs the trial against a model of the site built from its own telemetry—line speeds, cycle times, quality data—joined to R&D and demand data. The World Economic Forum's Global Lighthouse Network reports that P&G's Takasaki plant, which deployed data flow integration, digital twins, and machine learning across the value chain from R&D to customers, shortened innovation lead times by 72% and reduced shutdown days for trials by 21% (World Economic Forum, 2023); at its Lima site, digital twins and advanced analytics helped bring new products to market ten times faster (World Economic Forum, 2021). What makes this OC rather than a purchased tool is that the twin is only as good as the plant history it encodes. It is built out of operations and is worth little to anyone with a different production setup.

The AI jobs we observe at P&G are the labor that performs this conversion, and they cluster into the categories our classification identifies as most common at the firm:

- *Manufacturing & process systems—turning plant experience into simulatable models.* The digital twin work sits here. A platform head describes having "[s]pearheaded the development of an AI engine to optimize production lines" and "[o]wned the development and deployment of digital twin models for a new manufacturing platform." An IT director for industrial IoT describes "[d]riving P&G's companywide IIoT platform for managing near real-time ML manufacturing models." A modeling director describes a program "that aims at digitizing end to end the innovation process from an idea on the paper to the production line."

- *Supply-chain & logistics systems—embedding planning judgment into standing systems.* A supply chain executive describes deploying "AI-based route optimization tools (P&G ORION system)" and later leading "the development of a machine learning-based dynamic replenishment model," a framework subsequently adopted into a firmwide AI inventory platform. An engineering manager describes co-founding an "internal startup delivering AI/algorithmic warehouse optimization solutions across P&G's global network." A principal data scientist extends the twin idea from machines to processes, listing "[r]eal-time Digital Twins of business processes" including a "[c]ustomer order management digital twin to minimize on-shelf out-of-stocks."

- *Marketing & revenue-operations systems—consolidating consumer learning into reusable assets.* A product manager leads a Voice of Consumer suite described as "an enterprise ecosystem of AI, ML, and large language model (LLM) products that transform how 10 global businesses understand and act on consumer feedback." A brand director describes "[s]pearheading the use of generative AI in brand building, from

developing proprietary knowledge to conducting pilot projects and scaling successful initiatives."

- *Proprietary-knowledge GenAI/NLP systems—codifying tacit expertise directly.* A principal data scientist describes "Corporate Knowledge Management using Generative AI," comprising an "[i]nsight engine for explicit knowledge using large language models" alongside "[c]apture and insight building from tacit knowledge residing with Experts (SpeechBrain)"—an unusually explicit statement of the conversion this section is about, and one that names the instrument: a speech toolkit, applied to what experts say rather than to what they have already written down. The same description places this work directly alongside "[r]eal-time Digital Twins of business processes," so the codification of expert judgment and the codification of process behavior are treated as one program. In HR, a manager describes prompt-engineering a system to write job descriptions and co-developing a review-process assistant that became "the most used P&G-wide GPT assistant."
- *Reusable data platforms & pipelines—turning the firm's own data into shared, reusable assets.* A chief data and AI officer describes having "[b]uilt the enterprise AI Factory & GenAI platform, compressing time-to-value from months to weeks." A group scientist describes delivering "enterprise AI platforms integrating multi-agent architectures, workflow orchestration, and retrieval-augmented generation (RAG) ... powered by proprietary data."

Figure B.2 places these examples in the context of the firm's hiring. Panel (a) shows that OC-building AI roles were flat as a share of P&G's US positions until about 2013 and then rose sharply, roughly a tenfold increase by the early 2020s before leveling off, while other AI roles grew far more modestly and plateaued after 2020. The gap between the two series widens substantially over the period, so the firm's AI hiring has been directed overwhelmingly toward work that converts proprietary operations into durable assets rather than toward generic technical capability. Panel (b) breaks that series down into the types of work that qualify as OC-building. Reusable data platforms and pipelines account for much of the level throughout, and marketing and revenue-operations systems form a steady base—consistent with the consumer-knowledge and brand examples above—while supply-chain and logistics systems and cross-functional operating routines together reflect the planning and replenishment work described earlier. The timing of the takeoff is itself informative: the digital twin work predates the generative-AI wave by nearly a decade, since codifying plant telemetry into a simulatable model required only conventional machine learning applied to data the plants were already producing. That is why P&G's series accelerates from around 2014, whereas at Morgan Stanley it accelerates after 2021—document and knowledge work became codifiable at scale only once language models arrived. Consistent with this, the category of systems built on proprietary knowledge using generative AI and natural language processing is essentially absent before the early 2020s here too, arriving late but on top of a much larger pre-existing base.

The productivity channel here is unusually visible. A physical trial consumes capacity: the line stops, and output is forgone to learn something. Running the trial against a digital twin decouples learning from production, so the plant keeps producing while the

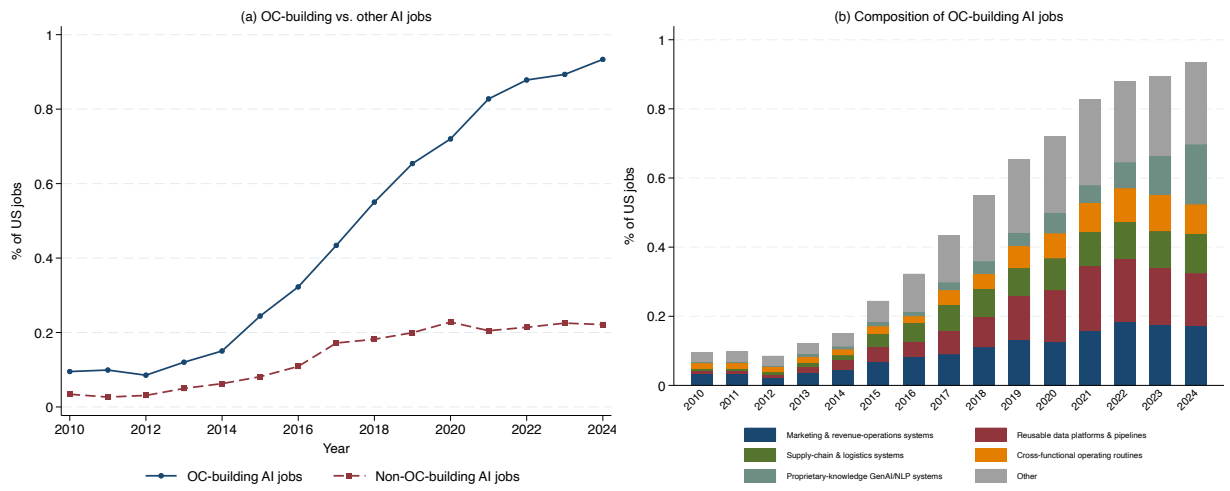


Figure B.2. AI jobs at Procter & Gamble as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification.

experiment happens elsewhere—a 21% reduction in trial shutdown days is additional output from the identical stock of plant and equipment. The same mechanism compresses the innovation cycle, since more candidate formulations and process settings can be evaluated per unit of engineering and plant time, which is what a 72% reduction in innovation lead time and a tenfold acceleration in speed to market represent. On the supply side, replenishment and routing models deliver the same service levels with less inventory and fewer miles, lowering the working capital and logistics cost embedded in each unit sold. And because a twin built at one site can be adapted to the next, the fixed cost of codifying is spread across a network of plants and lines rather than borne once per site.

Two features stand out. First, in nearly every case the input is data the firm was already generating—process telemetry, order flows, consumer feedback, supplier prices—and the output—a model of the plant, held in the firm's own systems—is an asset that does not depend on the engineers who built it. P&G's own account of its AI strategy emphasizes exactly this, unlocking insight from data the company already holds (Davenport and Bean, 2025). Second, a substantial share of this work is not any single application but infrastructure for producing applications: platforms, model governance, and internal accelerators that lower the cost of the next conversion. That is what makes the accumulation compound, and it is also why the payoff is unlikely to be immediate.

B.3.3. Moderna

Moderna—a pharmaceutical company that produced a COVID-19 vaccine in record time using its "AI factory" (Barbaschow, 2021)—presents a different starting point from that of the other two cases: a firm founded in 2010 and without a commercial product until 2020, which scaled from clinical-stage research to global manufacturing in a matter of years. Its job descriptions read accordingly: a computational sciences executive describes having

"[f]ounded, built, and led the first ever ML/AI capability at Moderna, the first custom informatics platform"; a commercial data director describes leading "the establishment of data infrastructure from the ground up for a rapidly scaling pharmaceutical company, transforming it from a startup to a commercial enterprise." What Moderna lacked was not operational knowledge but codified organizational assets—the procedures, routines, and reusable systems that a firm normally accumulates over decades. At the same time, the company describes a decade-long library of data as its single biggest asset, and being cloud-native from the outset as what let it avoid the data silos that slow codification elsewhere (Challame, 2023). A firm generating knowledge far faster than it has codified it is exactly where the mechanism we describe binds hardest.

Moderna's response was to make codification something every employee does rather than a project run by a central team. Its internal chatbot, mChat, was built in about two weeks and launched in May 2023; within months roughly two-thirds of employees were active users, embedding it into their own workflows (Challame, 2023). The company subsequently deployed ChatGPT Enterprise with a builder that lets employees create assistants trained on their own team's documents and terminology, and within roughly two months employees had built more than 750 such tools across legal, research, manufacturing, and commercial functions. The significance is not the count but the mechanism: the people who hold the operational knowledge are the ones converting it into durable form, without waiting for engineering capacity.

The AI jobs we observe at Moderna show what this looks like in practice, and cluster into the categories our classification identifies as most common at the firm:

- *R&D & scientific platforms—turning a short experimental history into predictive models.* The dominant category in Moderna's early years. An RNA process engineer describes having "founded and launched the RNA AI initiative by building predictive ML models from thousands of development experiment data points," work that "accelerated development time by several months, reduced unexpected manufacturing failures." A limited stock of experiments becomes a reusable asset that substitutes for experience the firm has not yet had time to accumulate. A quantitative biology executive describes integrating "robotic plate handling and AI/ML tools into NGS Prep" and improving "time to result from 4 hr/run to 20 min/run with improved accuracy."

- *Reusable data platforms & pipelines—building the firm's data foundation from nothing.* A VP of data and robotics platforms describes having "[d]esigned Moderna's data platform strategy" and "[d]eveloped and launched an AI platform integrating LLMs with proprietary data, both structured and unstructured," including an operational control tower for the supply chain. A VP of AI products describes introducing "AI as a utility to all knowledge workers through a company-wide transformation program."

- *Proprietary-knowledge GenAI/NLP systems—codifying process and documentation directly.* A manager describes developing "custom GPT tools, including 'Meeting Minutes GPT' and 'SOP Creation GPT,' significantly improving operational efficiency and reducing document preparation time from days to hours." An HR director describes

deploying "custom GPT solutions that reduced manual inquiries and strengthened knowledge management." Standard operating procedures are OC in its most literal form, and here they are written with the assistance of the tool.

- *Cross-functional operating routines—institutionalizing operating routines and coordination mechanisms.* A supply chain associate director describes championing "AI-driven process optimization using Mimica task mining to identify automation opportunities across planning and logistics workflows," and co-designing a "multi-agent strategy integrating AI, IoT, and digital twins to enable autonomous operational decision-making"—instrumenting operations to find what is worth codifying next.
- *Manufacturing & process systems—converting production and quality experience into reusable routines.* A quality specialist describes spearheading "AI-enabled, LEAN-based opportunity engines across Quality, CMC, and Manufacturing," extending codification into the regulated production environment the company had to build from nothing.
- *Hiring, training & talent development—institutionalizing AI skill-building into training curricula.* A talent development head lists the "launch of Moderna's AI Academy"; a learning director describes creating "Moderna's Digital College" with "comprehensive learning paths in AI, Cybersecurity, and Digital Tooling." This category is small in the aggregate but central here: distributing the capability to codify was the company's explicit strategy, so part of what the firm built here is the capability of the people doing the converting.

Figure B.3 places this in the context of the firm's hiring, and the levels are striking: OC-building AI roles run at one to three percent of Moderna's US positions, several times the shares at Morgan Stanley or P&G. Panel (a) shows the series dipping through 2020—the pandemic years, when the company was hiring at scale for manufacturing and distribution—and then roughly tripling between 2020 and 2024, while other AI roles stay flat below half a percent throughout. The acceleration begins after the commercial scale-up rather than during it, consistent with a firm that first built the operation and then codified it. Panel (b) shows what changed. R&D and scientific platforms dominate the early years, when AI at Moderna was largely a research capability; from 2021 onward the growth comes from reusable data platforms and pipelines, cross-functional operating routines, and manufacturing and process systems—the codification of how the company runs rather than what it discovers. Systems built on proprietary knowledge using generative AI appear from 2022 and expand quickly, matching the mChat and custom-GPT timeline described above.

The productivity channel is one the company states directly: whereas conventional operating models require headcount to grow with the business, its aim is to "scale the value of their people"—which is output per worker (Challame, 2023). The mechanism is visible in the individual cases. Predictive models built from a few thousand development experiments let process engineers avoid failures they have not personally encountered, so output rises not because inputs increased but because the firm stopped rediscovering the same things. Drafting SOPs, meeting records, and regulatory documentation consumes

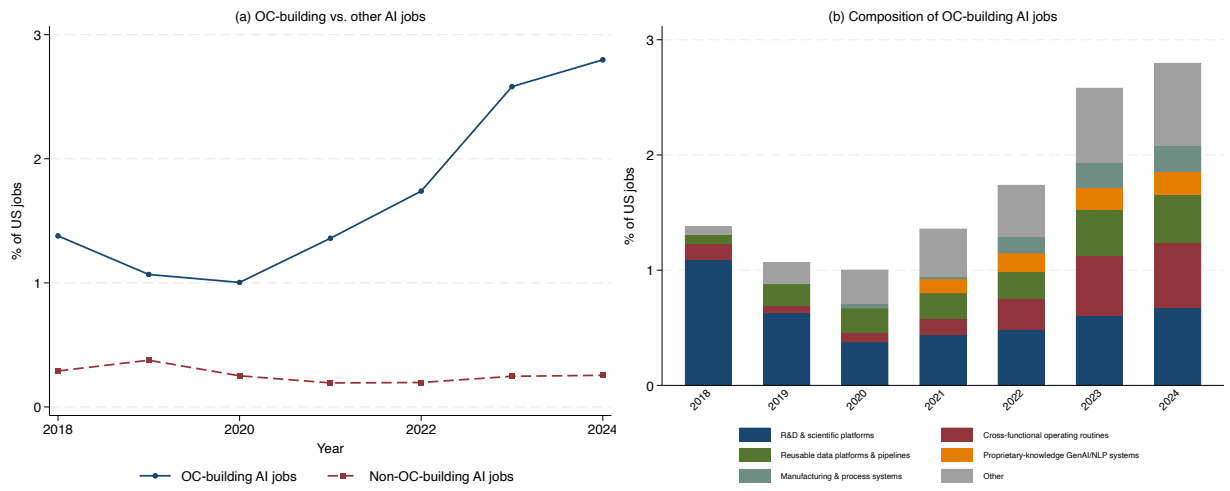


Figure B.3. AI jobs at Moderna as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification. Note the vertical scale differs from the corresponding figures for Morgan Stanley and Procter & Gamble.

scientific and professional hours that generate no output themselves; compressing that work reallocates skilled time toward activities that do. And because a GPT built by one team is available to the next at essentially no cost, each conversion raises the productivity of people who had no part in building it.

The broader point is the one our heterogeneity result makes. Moderna entered the AI era holding little OC as a new firm, and so had the most to gain from a technology that speeds the accumulation of OC—which is exactly what followed, as the company built that stock far faster than a firm accumulating it through ordinary experience could have. That it chose to distribute the capability to every employee rather than centralize it follows from the same reasoning: when the codified stock is thin relative to the operation, the binding constraint is the rate of conversion, not the quality of any single converted asset.

B.3.4. Home Depot

Home Depot presents the mechanism in its most literal form. The firm's competitive asset has always been what it calls orange apron knowledge—the expertise of store associates, many of them former tradespeople, who can tell a customer which grout works with glass tile and where to find it. That knowledge is the definition of tacit: it lives in individuals, it is available only in one aisle of one store at a time, and it walks out the door with the associate. In a retail workforce of over 470,000 associates across more than 2,300 stores, with the turnover rates the sector implies, the firm has historically had to rebuild it continuously. AI lets Home Depot convert more of that knowledge into a stock the company holds.

The clearest instance is Magic Apron, an assistant that began on product pages and has become a conversational tool spanning the firm's digital platforms, extending to real-time store inventory and aisle-level product locations. What it is built from is the point. The

company describes it as powered by a proprietary home improvement knowledge base that combines large-scale datasets with its own expertise and product information, trained on Home Depot's proprietary project expertise and product catalog alongside information from the wider internet, and explicitly designed to leverage the company's own human expertise; the tool continues to learn from customer and associate feedback and serves as a foundation for store and contact center associates (The Home Depot, 2025a,b). The executive who leads the effort has framed the conversion in both directions—customers have always relied on the expertise of orange-aproned associates in the aisles, and the aim is to bring that same expertise to the digital world, "putting 'Orange Apron' expertise in the pocket of every customer" (The Home Depot, 2026b). Alongside it sit AI-powered materials lists, which let pros describe a job in plain language and receive a grouped list including items they had omitted, and blueprint takeoffs that produce quotes in days rather than weeks (The Home Depot, 2026a). Less visible but equally relevant are the operational systems: order intelligence that identifies delivery misses before they reach the customer, and route intelligence that layers customer-specific data with road and weather conditions to anticipate access problems at the jobsite (The Home Depot, 2026b).

The AI jobs we observe at Home Depot show the work of conversion, and cluster into the categories our classification identifies as most common at the firm:

- *Reusable data platforms & pipelines—assembling orange apron expertise into a queryable stock.* The largest category. The knowledge base behind Magic Apron is built from the company's accumulated project expertise and refined continuously by associate feedback, and the job descriptions show the internal counterpart: a senior director of data science describes leading enterprise generative AI foundation work to "build the federated home improvement knowledge brain and AI common services" for the company's applications.
- *Firm-specific predictive & decision models—embedding store-level experience into forecasting and decision models.* A principal data and analytics lead describes engineering "a novel On-Shelf Availability (OSA) metric by integrating inventory, bay sequencing, and pack-down data"—converting what experienced staff knew informally about stock position into a predictive measure usable everywhere. A senior director of operations describes adopting "AI and machine learning to improve labor forecasting, task prioritization, and operational visibility."
- *Marketing & revenue-operations systems—turning customer interactions into structured project knowledge.* A senior data science manager describes having "[u]sed deep learning & knowledge graph to recognize customer's DIY project intents in real time, providing holistic product, how-to guide, rental and home services recommendations"—customer interactions turned into structured project knowledge.
- *Reusable software platforms & workflows—standardizing execution across a distributed store network.* A product leader in operations describes designing "platform capabilities that strengthened visibility, standardization, and performance consistency across a distributed store network," with tools that "empowered frontline associates with real-time insights." A manager in inventory transformation describes launching

"a reset platform, anticipated to eliminate 450 hours per week of non-value-added work."

- *Cross-functional operating routines—codifying frontline procedure.* An analyst in process and knowledge management describes "simplifying process and instruction documentation using Lean Six Sigma (LSS) and Knowledge-Centered Service (KCS) methodologies," work that improved guided workflow integration while "reducing onboarding costs by $3k per hire"—documentation as the mechanism by which a new associate reaches proficiency faster. A conversational UX designer describes "[e]stablishing Conversation Design practices at the Home Depot" and creating a guild to teach others where conversational AI applies.

Figure B.4 places these examples in the context of the firm's hiring. Panel (a) shows OC-building AI roles rising almost continuously as a share of Home Depot's US positions, with the increase steepening after 2014 and continuing through 2024, while other AI roles grow more slowly and flatten after 2018. The levels are worth noting: they are well below those at the technology-intensive firms in our set, which is what one expects of a retailer whose workforce is overwhelmingly in stores rather than at headquarters. Panel (b) decomposes that series. Reusable data platforms and pipelines and marketing and revenue-operations systems form much of the base, consistent with the consumer-knowledge and merchandising work described earlier, while firm-specific predictive and decision models together with supply-chain and logistics systems reflect the on-shelf availability, labor forecasting, and delivery examples. The composition is notably stable over time relative to the other firms we examine: Home Depot was codifying store and supply-chain operations with conventional machine learning long before generative models arrived, and the recent Magic Apron work sits on top of that accumulated base rather than initiating it.

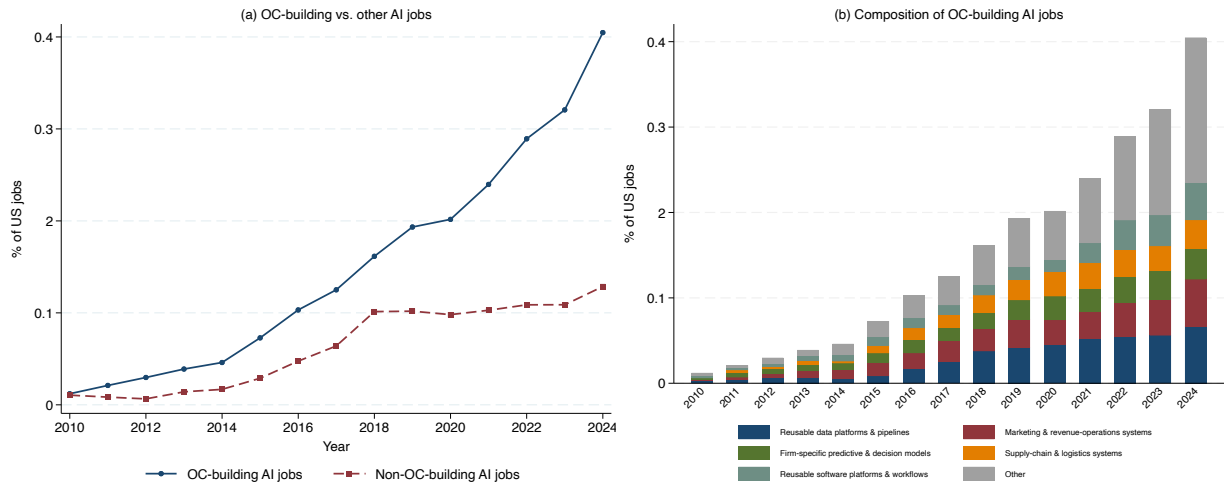


Figure B.4. AI jobs at Home Depot as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification.

The productivity channel is sharp here because the multiplication factor is large. Expertise codified once is available in every aisle of every store simultaneously, at no additional cost per store; the same asset that helps one associate helps all of them. This matters most precisely where embodied knowledge is least durable. In a workforce with high turnover, an associate's departure means losing what they knew, and the firm pays to rebuild it with each hire—which is why the onboarding-cost reduction above is not a minor administrative saving but the mechanism working. On the customer-facing side, materials lists and blueprint takeoffs compress estimating work that generates no output itself, and order and route intelligence prevent failed deliveries, which are pure waste of labor, fuel, and truck capacity.

B.3.5. CVS Health

CVS Health presents the organization capital mechanism in an unusually integrated setting. The company spans roughly 9,000 retail pharmacy locations, more than a thousand walk-in and primary care clinics, a pharmacy benefits manager with about 88 million plan members, and health insurance serving more than 37 million people through Aetna (CVS Health, 2026a). Operational knowledge accumulates across all of it automatically, as a by-product of filling prescriptions, paying claims, and managing care. The constraint has never been generating that knowledge but rendering it usable: much of it arrives as unstructured clinical documentation—as Aetna's chief digital and technology officer describes the records, "medical charts that have white space filled with handwritten notes" (Olavsrud, 2026). The firm possessed this knowledge all along without being able to act on it at scale.

The clearest instance is AI Medical Chart Review at Aetna. Under HEDIS quality measures, large managed care providers review more than ten million medical records annually to identify gaps in care—missed screenings, overdue tests, lapsed chronic disease management. Industry benchmarks put the annual review at roughly 50,000 work weeks, nearly a thousand dedicated full-time employees; a team of fifty working manually would need nearly two decades to complete a single year. The platform extracts clinically relevant data and ranks records by likelihood of measure closure and evidence strength, passing a prioritized set to human reviewers. It has cut the need for manual review by 65% and processed fourteen million documents, with capacity formerly spent on review shifting toward quality control (Olavsrud, 2026). The same logic runs through claims and authorization: the second-generation Claims Assist Manager cuts processing time by more than 20% on complex claims requiring manual review (CVS Health, 2026a), and prior authorization reform has standardized data requirements for 88% of volume, moved 83% of authorizations to real-time processing, and eliminated more than a million provider calls (CVS Health, 2026b). On the pharmacy side, member-facing and colleague-facing tools draw on the same accumulating record, and CVS Health has since announced Health100, a subsidiary intended to deliver an integrated engagement platform across these surfaces (CVS Health, 2026b).

The AI jobs we observe show the work of conversion, span the company's businesses, and cluster into the categories our classification identifies as most common at the firm:

- *Reusable data platforms & pipelines—making a record the firm already held usable.* By

far the largest category, and the binding constraint for a firm whose problem is not generating knowledge but rendering it usable. A lead director of machine learning engineering describes "[b]uilding the world's largest RAG system for knowledge management for CVS Health's 300,000+ employees that will scale to 1M+ documents." A senior manager in data governance describes pioneering "creation of a business glossary, standardizing terminology" across the firm's data.

- *Proprietary-knowledge GenAI/NLP systems—extracting clinical knowledge from unstructured documents.* An NLP technical lead describes responsibility for "the creation and maintenance of the enterprise NLP platform," used across domains for "bot creation, customer IVR systems, financial contract extraction, and legal text analysis." An AI research architect describes having "mapped over 1,600 distinct healthcare outcomes and their associated friction points," work that "collapsed delivery timelines from six weeks to one," and a research operations manager describes an "AI-enhanced research repository with automated tagging and intelligent retrieval."
- *Cross-functional operating routines—codifying procedure as the precondition for AI.* An executive director of employee experience describes having "[d]igitized and standardized HR knowledge, policy, and procedure content and implemented a governance framework as the foundation for AI"—an explicit statement that documentation comes first and the model second. A head of enterprise AI governance describes establishing the company's Responsible AI operating model and policy framework across all lines of business, necessary infrastructure in a regulated setting and codified organizational practice in its own right.
- *Reusable software platforms & workflows—translating pharmacy practice into standardized workflows.* A director of product management describes leading development of "an AI-powered platform that helped reduce medication errors across 9,000+ pharmacies." A platform head describes having "[b]uilt and scaled a Virtual Pharmacist to answer drug and side-effect questions," and a retail IT director describes building "StoreOS, a unified IoT/AI-driven platform redefining colleague workflows" including pharmacy pill verification AI.
- *Security, compliance & risk systems—turning audit and control experience into full-coverage review.* A director of business innovation describes having "[r]eplaced manual sampling with algorithm-driven 100% transaction review, reducing compliance risk," alongside unlocking eighteen thousand workforce hours through automation. In the PBM, a lead director describes delivering a "GenAI-powered Audit Advisor that reduced contract risk and improved rebate accuracy."

Figure B.5 places these examples in the context of the firm's hiring. Panel (a) shows OC-building AI roles growing slowly through the early 2010s, accelerating sharply after 2017, and continuing to climb steeply through 2024, reaching a level roughly three times that of other AI roles. The two series diverge progressively rather than moving together, so the firm's AI hiring has been directed increasingly toward work that converts proprietary operations into durable assets. Panel (b) shows what that work consists of, and the composition is distinctive among our cases: reusable data platforms and pipelines dominate

throughout and account for close to half the total by the end of the sample. That is what one expects of a firm whose codification problem is less about capturing knowledge that was never recorded than about making usable an operational record it already possessed—the pipelines are the binding constraint. Cross-functional operating routines and marketing and revenue-operations systems form steady secondary components, and systems built on proprietary knowledge using generative AI and natural language processing appear from the early 2020s and expand quickly, matching the chart review, claims, and conversational tools described above.

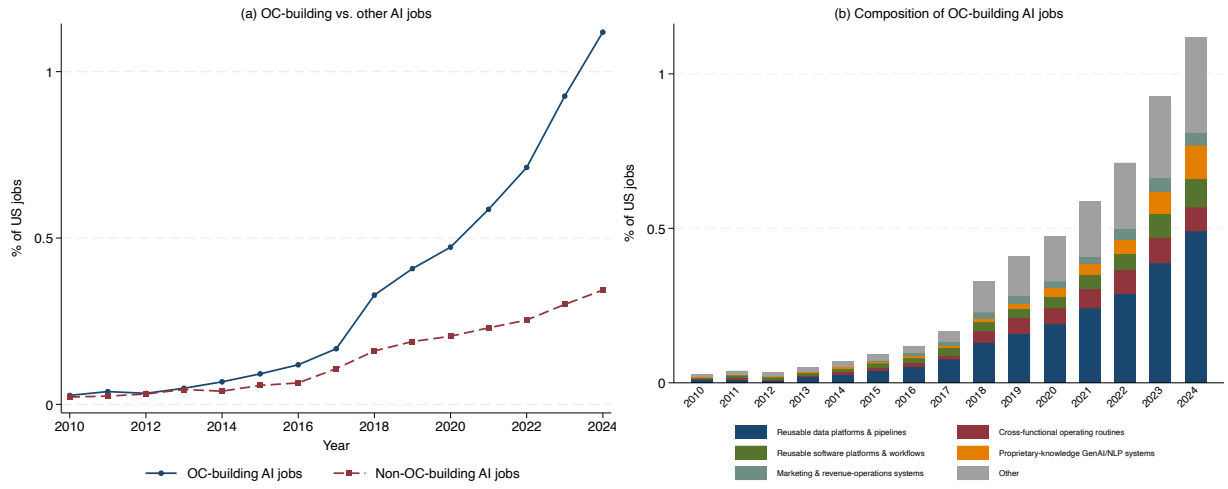


Figure B.5. AI jobs at CVS Health as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification. Note the vertical scale differs from the corresponding figures for the other firms.

Two productivity channels stand out, and the first is distinctive to this case. In our other examples AI speeds up work the firm was already doing; here it makes feasible a task that—without AI—is not fully possible. Reviewing every chart, or every transaction rather than a sample, was never a matter of working faster—it exceeded what any realistic workforce could do, which is why the industry relied on partial review. Codification converts sampling into census, and the firm learns things about its own operations that were previously unknowable at any cost. The measured gains follow: the same coders process a far larger share of the record, with freed capacity redirected to quality control rather than eliminated, so output per worker rises without headcount falling; higher gap closure feeds Star Ratings and reimbursement, making this a revenue channel and not only a cost one; and eliminating a million provider calls removes administrative work from both sides of a transaction. The second channel is integration. Because the same person may fill a prescription at a CVS pharmacy, see a clinician at MinuteClinic, and file a claim through Aetna, knowledge codified in one business is usable in the others—an asset built once and drawn on repeatedly across surfaces that a monoline insurer or a standalone pharmacy could not connect.

B.3.6. Verizon

Verizon runs two large workforces whose knowledge has historically been hard to hold: tens of thousands of field technicians who install and repair the network, and tens of thousands of care representatives who resolve customer problems. In both cases the expertise is procedural, learned on the job, unevenly distributed, and lost when someone leaves. The firm also generates an enormous operational record—dispatch logs, trouble tickets, network telemetry, support transcripts—as a by-product of running the network. AI lets it convert more of that record into an asset the company holds.

The clearest instance is on the care side, and its construction is explicit about what is being converted. Verizon's internal assistant was built by feeding roughly fifteen thousand internal documents into an LLM (Cai, 2025). Deployed across 28,000 care representatives and retail locations, the Personal Research Assistant surfaces context-aware answers and has been extended with automated conversation summarization and follow-up reminders; a companion Problem Solver agent has been reported to benefit newer representatives in particular, raising their confidence on complex queries. The reported effect was not substitution but reallocation: shorter calls freed representatives to sell, with sales through the service team up sharply, which Verizon's consumer chief described as reskilling care agents into selling agents (Cai, 2025). On the network side, the company has moved toward agentic automation running on its own cloud platform (Benton Institute for Broadband and Society, 2026).

The AI jobs we observe show the work of conversion, and cluster into the categories our classification identifies as most common at the firm:

- *Customer-service & support systems—putting the best representative's answer on every screen.* The largest category at Verizon by a wide margin, and the one no other firm in our set shares. A head of data and analytics describes "AI-enabled productivity platforms" transforming "workforce operations across 65K+ retail and call center employees." The internal assistant built on the firm's own documents belongs here, and its effect was reported as reallocated time rather than reduced headcount.

- *Reusable cloud, infrastructure & reliability operations—codifying network and field operating procedures into automation.* A senior director of software development describes driving "Zero-Touch Operations (ZTO) across Verizon's wireless and wireline network operations" and pioneering a "Dispatch and Field Operations platform, achieving 85% digitization of manual service transactions via Machine Learning." A network operations technology head describes systems enabling a "30k+ workforce to be efficiently dispatched," delivering "mobile tools for our frontline 30k+ customer service technicians and their supervisors." An associate fellow describes an "OpenSON ecosystem," a "real-time, event driven automation ecosystem employing AI/ML pipelines" across radio access network domains.

- *Reusable data platforms & pipelines—unifying the record so it can be codified.* A head of data and analytics describes designing an organization "integrating data engineering, science, and governance under a unified model," the precondition for codifying anything at scale across a firm of this size.

- *Cross-functional operating routines—institutionalizing coaching as a standing management routine.* A transformation manager describes leading development of "Smart AI Coach (SAIC), an early AI application that provided technicians and their managers with real-time comparative results and coaching information," and a Data Driven Process Analysis program deployed after acquiring "data feeds from over a dozen separate systems" and training "over 1600 Local Managers." A vice president for data and AI describes establishing an enterprise AI council and Responsible AI practice across care, marketing, operations, and network.
- *Knowledge organization & ontologies—organizing specialized knowledge into ontologies and semantic layers.* Small in the aggregate but unusually explicit here. A principal engineer describes having "[b]uilt the Knowledge Engineering discipline from scratch, authoring lifecycle standards and governance frameworks adopted across Data & Analytics," and designing "a Unified Semantic Knowledge Layer with 1800+ classes supporting data literacy through business-friendly querying," including a "Network Operations ontology enabling semantic search." Few descriptions in our data state the mechanism as directly.

Figure B.6 places these examples in the context of the firm's hiring. Panel (a) shows OC-building AI roles rising steadily from 2010 and accelerating from about 2015, reaching a level roughly three times that of other AI roles by 2024, with the gap widening throughout rather than converging. Panel (b) explains why this case looks different from the others in our set: customer-service and support systems are the single largest category in every year and account for close to a third of the total by the end of the sample. No other firm we examine has its codification effort so concentrated on a single function. That concentration matches the case: the binding constraint is the dispersion of procedural knowledge across tens of thousands of care representatives and field technicians, so the asset worth building is the one that makes the best representative's answer available to all of them. Reusable data platforms and pipelines and cloud and reliability operations form the infrastructure beneath it, cross-functional operating routines appear steadily from the mid-2010s, and systems built on proprietary knowledge using generative AI and natural language processing emerge from the early 2020s—the timing of the assistant built on the firm's own internal documents.

Two features of the productivity channel stand out. First, the gains land disproportionately on the less experienced. The care assistant compresses the gap between a new representative and a seasoned one by making the seasoned one's answers available on screen, which is what codifying dispersed expertise into a common asset should do, and it is why the reported effect appears as reallocated time rather than reduced headcount—shorter calls converted into selling time. Second, the field organization illustrates the same logic in physical work. Digitizing manual service transactions and coaching technicians against comparative results turns the tacit skill of the best installers into a standard the rest can follow; the output is more jobs completed correctly per technician-day, with fewer repeat visits. In both cases the asset is non-rival within the firm: an answer codified once serves all 28,000 representatives, and a dispatch improvement serves the whole footprint.

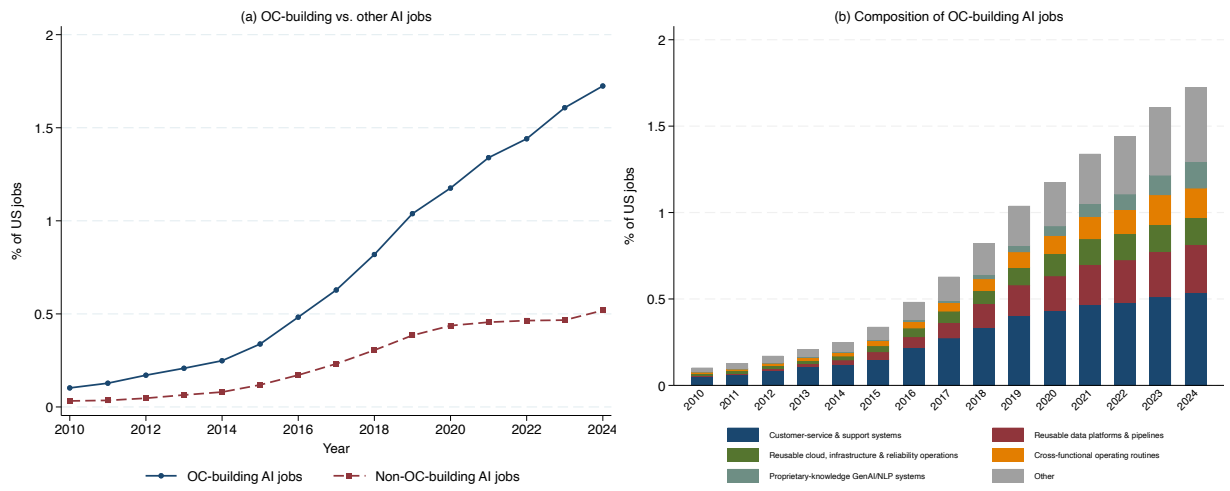


Figure B.6. AI jobs at Verizon as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification. Note the vertical scale differs from the corresponding figures for the other firms.

B.3.7. Nike

Nike presents a variant of the mechanism our other cases do not. Much of the operational knowledge the firm depends on is generated outside its own walls: Nike designs and sells footwear and apparel but does not own the factories that make it. The knowledge of how a shoe is actually built—tooling, materials behavior, line balance, quality yield, cost drivers—accumulates in a network of contract manufacturers and in the experience of the Nike staff who work with them. That knowledge has historically been partly tacit and partly stranded at the partner, which makes codification both more valuable and harder than in a firm that owns its plants. Nike acquired several analytics and data-integration firms between 2018 and 2021 (Kitishian, 2025; Sohel, 2026), but the acquisitions supplied components rather than the asset itself; what the internal roles below describe is the work of attaching those components to the firm's own processes and records, which is where the stock is actually built.

The clearest internal instance is the manufacturing backbone. A senior director describes being "[p]romoted to jump-start and lead the first digital transformation of Nike's manufacturing landscape of 500+ factories, through modern platforms, software and connectivity," producing "a connected and resilient ecosystem of multiple Nike functions and their manufacturing partners, delivering 1.4B+ units of Nike product on time." The instrument is described as "the industry leading data-streaming and partner-collaboration Nike Manufacturing Backbone platform," digitizing "critical supply chain collaboration, visibility, resilience, and sustainability business processes." The object being built is a shared record across a boundary the firm does not control.

The AI roles we observe show the work of conversion, and cluster into the categories our classification identifies as most common at the firm:

- *Reusable data platforms & pipelines—merging decades of incompatible records into one*

source of truth. The largest category, and the step at which dispersed operational records become a single addressable stock. A director of enterprise data and analytics describes building a cloud data lake as "one 'source of truth' for data and insights," ingesting "transactional ERP, point of sale, digital/consumer, and product creation" data. The records being merged are the firm's own, accumulated over decades in incompatible systems.

- *Product & design systems—codifying product-creation expertise into tools.* A director of footwear cost engineering describes shifting the function "from 18 Individual Contributors to a 4-person Center of Excellence (COE) that provides 'smart' tools to Product Creation team, enabling quick decisions," including "an AI/ML-based auto-budgeting tool that reduced creation time from 3 hours to 2 minutes"—costing judgment that had lived in eighteen experienced people becoming a tool available to the whole creation organization. The publicly documented design-side analogue is the same substitution at larger scale: Nike's Athlete Imagined Revolution project applies generative models to footwear concept generation with elite athletes, compressing concept-to-prototype from weeks or months to hours (NIKE, Inc., 2024; Team DigitalDefynd, 2026), with the underlying asset described as a proprietary model trained on decades of Nike's own athlete performance data rather than a general corpus (Kitishian, 2025).

- *Supply-chain & logistics systems—codifying supplier and fulfillment learning across factories Nike does not own.* A senior director describes owning product and platform strategy for "Manufacturing, Digital Consumer Order Management, Wholesale Customer Order Management, Inventory Management, Global Fulfillment, Global Distribution, Global Customs & Trade, & Global Transportation," directing hundreds of engineers across a portfolio intended to integrate these processes end to end. The manufacturing backbone described above is the flagship instance of this category.

- *Cross-functional operating routines—agreeing what the numbers mean across functions.* An enterprise data director describes pairing the data lake with "governance, master data management," and "KPI/Metric alignment across stakeholder teams"—agreeing what a metric means across functions is itself codified practice, and the precondition for the marketing and revenue-operations work that forms the demand-side counterpart to the supply-chain systems above.

- *Reusable AI/ML platforms & MLOps—turning model-development experience into tooling many teams can use.* A vice president for global architecture describes leading the transformation of "a $47B company with over 79,000 employees, and over 50 years of technology accumulation, through a platform transformation." A senior director of technical product management describes an ecosystem of "standardized, fully managed and composable control planes" spanning developer, data, and AI capabilities, exposed to agents through standard interfaces.

Figure B.7 places these examples in the context of the firm's hiring. Panel (a) shows OC-building AI roles essentially flat through 2013 and then rising sharply, with a particularly steep step around 2018 and continued growth to roughly five times the level of

other AI roles by 2024. The timing is consistent with the case: the manufacturing and supply-chain digitization described above dates from the late 2010s, and the acceleration follows it rather than preceding it. Panel (b) decomposes the series. Reusable data platforms and pipelines form the base throughout, with marketing and revenue-operations systems and supply-chain and logistics systems as substantial and roughly comparable components—a split that matches a firm codifying on two fronts at once, consumer demand knowledge on one side and manufacturing and distribution knowledge on the other. Cross-functional operating routines and reusable AI/ML platform work grow steadily from the late 2010s, reflecting the platform and center-of-excellence roles described above. Notably, no generative-AI category appears in Nike's composition, in contrast to Morgan Stanley, CVS Health, and Verizon. Nike's codification effort runs through structured operational and commercial data rather than through documents and language, which is what one would expect when the knowledge being converted concerns how a product is costed, made, and moved rather than what an expert would say.

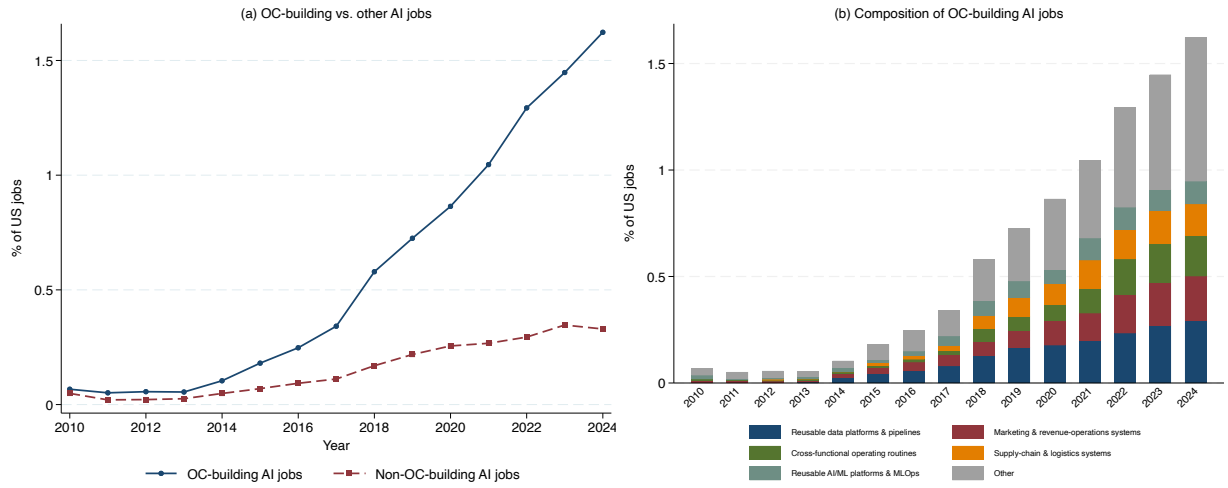


Figure B.7. AI jobs at Nike as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification. Note the vertical scale differs from the corresponding figures for the other firms.

The productivity channel runs through decision speed and through yield. Product creation at Nike is a repeated design problem under cost and performance constraints, and the binding input is the time of experienced designers, developers, and engineers. Converting costing and specification judgment into tools does not make those people better; it means the median decision is made with the knowledge of the best decision-maker, in minutes rather than hours, so more design iterations occur per unit of skilled time. On the manufacturing side, a shared digital record across hundreds of factories reduces the coordination failures—late changes, misread specifications, quality escapes—that consume material and rework capacity at the partner and delivery reliability at Nike. On the inventory side, the return is the difference between a demand forecast made once per season and one made continuously, which shows up as avoided markdowns rather

than as higher output.

One qualification follows. Knowledge codified into a partner-facing platform is not entirely firm-specific: contract manufacturers serve competing brands, so practice that Nike documents and teaches may diffuse. Our assumption that the codified stock stays with the firm holds better for some forms of the asset (e.g., internal documentation) than for others, and it is weakest for a shared platform spanning independent suppliers, which makes Nike the case where appropriability is most visibly at issue.

B.3.8. Expedia Group

Expedia Group presents the mechanism in a firm whose product is software and whose operational knowledge is therefore largely engineering knowledge: how its services fail, how incidents are diagnosed, how a feature moves from specification to production, how the marketplace behaves under stress. That knowledge accumulates in incident records, code repositories, runbooks, and the practice of individual engineers, and it has historically been unevenly distributed—the best engineer on a team knows things the others do not, and the firm holds only part of what they know apart from their continued presence. What makes this case distinctive is that Expedia has treated the diffusion of that expertise as an explicit strategy rather than a by-product.

The company's approach has two parts. The first is infrastructure: a shared internal stack built before any particular application, including agent testing and evaluation infrastructure and an internal environment giving employees access to many large language models to build their own assistants. Since early 2025 employees have built more than 1,500 agents in that environment (Fortune, 2026). The second is the diffusion strategy, and it states the mechanism of this paper almost exactly. After a nine-month study of how its roughly 5,000 engineers used AI coding assistants, the company found that its strongest performers gained far more than the average engineer, and set out to identify those engineers, capture their methods, and spread them across the organization (Lee, 2026). One engineer is described as having assembled a personal workbench of roughly ten agents supported by fifty to sixty specialized skills, changing his work from writing code to orchestrating agents (High, 2026). Codifying that workbench so others can use it is precisely the conversion of individual expertise into an asset the firm can hold and redeploy.

The AI jobs we observe show the work of conversion, and cluster into the categories our classification identifies as most common at the firm:

- *Reusable AI/ML platforms & MLOps—codifying how software gets built.* The shared stack the company built before pursuing particular applications. A director of software engineering describes leading development of a business-availability AIOps solution and having "[a]rchitected Agentic AI integrated with observability platforms, assisting developers with SRE principles and reducing support tickets through internal knowledge base ingestion," alongside "automated single-click workflows via Slack/UI, improving developer productivity by 30%." A director of product management for developer experience describes driving "AI-native development and tooling across the SDLC adopted by 80% of engineers, including an MCP registry,

agentic workflows, and spec-driven development templates," reporting a 20%+ improvement in developer productivity. Templates and registries are codified practice: they make one team's way of working the default for the next.

- *Reusable data platforms & pipelines—building the data substrate codification requires.* A senior director of machine learning engineering describes delivering "structured clickstream eventing, visitor context sharing, segmentation, fenced pricing, bot detection, identity classification, telemetry, knowledge graph, data discovery and federated governance," and unifying traveler identity across brands. A data stewardship director describes establishing "a data management program from the ground up" with a "hybrid" model pairing a central team with domain stewards across the organization.
- *Marketing & revenue-operations systems—making experimentation cheap enough to learn from.* The largest category in the firm's composition. A vice president for data platform and products describes "the consolidation of three brand-specific A/B testing platforms into one, resulting in a 5x increase in experimentation velocity"—infrastructure that converts commercial guesses into recorded results and improves future targeting and launches.
- *Firm-specific predictive & decision models—embedding marketplace learning into decision systems.* A technology vice president describes delivering "a world-class fraud detection decision engine" processing petabytes of data and "computing 100K+ complex features in milli-seconds, training and executing hundreds of models, and thousands of rules to make split-second risk decisions."
- *Reusable cloud, infrastructure & reliability operations—turning incident knowledge into an assistant.* A lead site reliability engineer describes leading "war rooms" for high-priority incidents and mentoring engineers on troubleshooting, then building an in-house capability that "replaced third-party limitations with a proprietary AIOps system capable of automated incident correlation and real-time revenue impact analysis," alongside a global "Gap Process" for around-the-clock operations pods.
- *Customer-service & support systems—codifying service knowledge.* A product vice president describes influencing the "transformation of the business unit from a traditional call center operation to an AI-powered conversational and self-service platform," delivering a multi-channel conversation platform.

Figure B.8 places these examples in the context of the firm's hiring, and the levels are the highest of the eight firms in our case-study sample—OC-building AI roles reach over four percent of Expedia's US positions by 2024. That is what one expects of a firm whose product is software and whose workforce is concentrated in engineering rather than in stores, plants, or branches. Panel (a) shows the series rising almost continuously from 2010 with no discernible inflection at the arrival of generative AI, reaching roughly two and a half times the level of other AI roles. Expedia has been codifying operational knowledge for the whole period; the recent agentic work extends a long trajectory rather than starting one. Panel (b) is consistent with the case. Reusable data platforms and pipelines and reusable AI/ML

platform and MLOps work together account for a large share throughout—the shared stack that the company built before pursuing particular applications—alongside firm-specific predictive and decision models and marketing and revenue-operations systems, reflecting the marketplace pricing, fraud, and traffic-acquisition work described above. Customer-service and support systems form a smaller but steady component, consistent with the conversation platform. As at Nike, no generative-AI category appears: Expedia's codification runs through platforms, models, and structured operational data rather than through documents and language.

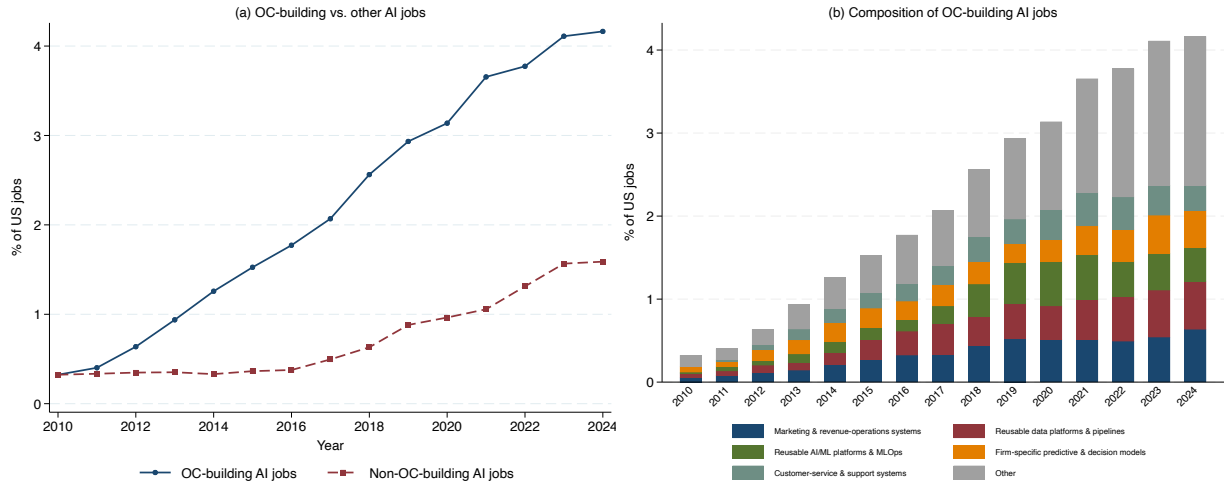
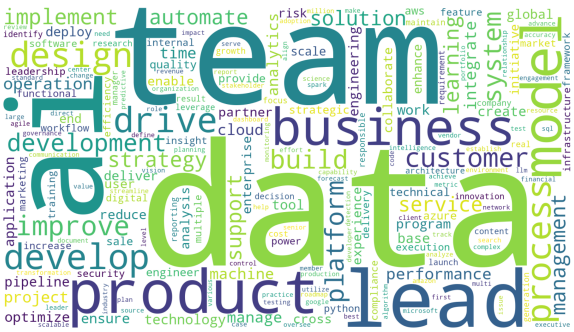


Figure B.8. AI jobs at Expedia Group as a share of the firm's US positions. Panel (a) compares OC-building AI jobs with non-OC-building AI jobs. Panel (b) decomposes the OC-building jobs by reason for classification. Note the vertical scale differs from the corresponding figures for the other firms.

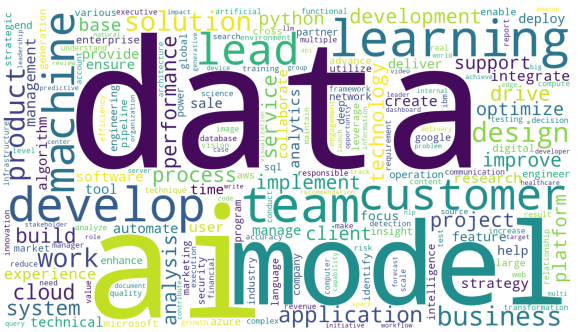
The productivity channel is well documented at Expedia, and it has a distinctive shape. Reported adoption of coding assistants rose from roughly two-thirds of developers with an estimated 20% productivity lift in early 2026 (Fortune, 2026) to 92% of engineers a few months later, with more than a tenth reported to be achieving two- to fivefold improvements (High, 2026). The dispersion is the interesting part. Where codified expertise compresses the gap between novice and expert in the cases we examine elsewhere—care representatives, store associates, underwriters—here the reported gains are concentrated among the already-strong-performing employees. Both patterns are the same mechanism seen from opposite ends: the value comes from making one person's method available to everyone, and whether that raises the floor or the ceiling depends on whose method is being captured. The company also frames the resulting capacity as substitutable for headcount at the margin, with its CTO suggesting that token spending may become a standard workforce budget line (Lee, 2026)—an unusually direct statement that codified capability and labor can serve as alternatives—a possibility we explore in an extension of the model in Section E.6.

Figure B.9. Word Clouds for Organization-Capital-Building and Non-Organization-Capital-Building Jobs

(a) AI, OC-building



(b) AI, Non-OC-building



(c) Non-AI, OC-building



(d) Non-AI, Non-OC-building



This figure plots word clouds for four job groups defined by the intersection of our AI-related job indicator and organization capital (OC) classification: (a) AI-related jobs that build organization capital (OC-building); (b) AI-related jobs that do not build organization capital (Non-OC-building); (c) non-AI jobs that build organization capital; and (d) non-AI jobs that do not build organization capital. The word clouds are constructed from job descriptions of U.S. jobs in the U.S. public-firm sample. For each job group, words are extracted from positions classified into that group based on our AI-related job indicator and LLM-based organization capital classification. Words are lemmatized and related derivational forms are merged. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure B.10. Average Organization-Capital Score: AI versus Non-AI Jobs



This figure plots the average continuous organization-capital (OC) score, which ranges from 0 to 1 and measures the extent to which a job builds OC, from 2010 to 2024 in the U.S. public-firm sample. Panel (a) plots the average OC score of all AI-related jobs. Panel (b) plots the same average score separately for tech firms (NAICS 51, 54, and semiconductor manufacturing) and all other firms. Panel (c) plots the average OC score of non-AI jobs, based on the random sample of 1,000 non-AI jobs per year. Panel (d) plots that average score separately for tech and non-tech firms. All averages are unweighted across jobs. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table B.3. Representativeness of the Non-AI Sample Used for Organization-Capital Scoring

Panel A: Distribution across industries (% of non-AI job-years)

Sector	Population %	Sample %
Manufacturing	27.02	27.83
Finance/Insurance	17.67	16.14
Retail Trade	14.64	14.60
Transportation/Warehousing	4.02	3.19
Accommodation & Food Svcs	3.67	3.87
Health Care	2.68	2.49
Wholesale	2.03	1.78
Utilities	1.71	1.39
Admin Support	1.65	1.86
Real Estate	1.42	1.11
Mining	1.25	1.19
Education Svcs	0.46	0.47
Arts/Entertainment	0.44	0.60
Construction	0.37	0.28
Other Svcs	0.26	0.21
Agriculture	0.07	0.07

Panel B: Firm size (US jobs per firm-year)

U.S. jobs per firm-year	Population	Sample
10th percentile	2,923	2,959
25th percentile	8,044	8,291
Median	23,457	24,207
75th percentile	70,311	73,298
90th percentile	153,629	154,255
Mean	52,280	53,728

This table compares the randomly drawn non-AI job sample with the population of non-AI U.S. jobs from which it is drawn. This sample serves as the non-AI benchmark in the organization capital (OC) analysis, and each sampled job is scored with the same OC classifier as the AI-related jobs. The population includes all non-AI U.S. jobs in firm-years with at least one AI-related U.S. job. The sample consists of 15,000 randomly drawn non-AI U.S. jobs from this population, with 1,000 jobs sampled per year from 2010 through 2024. Panel A reports the sector (based on 2-digit NAICS industry codes) distribution, excluding IT sectors. For each sector, the table reports the number of non-AI U.S. jobs in that sector divided by the corresponding total number of non-AI U.S. jobs, expressed as a percentage, separately for the population and the random sample. Panel B reports the distribution of firm size, measured as the total number of U.S. jobs at the firm-year level. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table B.4. Organization-Capital-Building AI versus Non-AI Jobs

	AI jobs	Non-AI jobs
N	593,332	15,000
Mean OC score	0.601	0.424
Share OC-building (score ≥ 0.6)	57.5%	31.0%
Distribution across score bands		
Unrelated (0.00–0.15)	2.6%	17.8%
Incidental (0.16–0.40)	21.3%	35.3%
Mixed (0.41–0.60)	18.6%	15.9%
Substantial (0.61–0.85)	48.7%	28.7%
Core (0.86–1.00)	8.7%	2.3%

This table compares organization capital (OC) scores for AI-related and non-AI U.S. jobs from 2010 to 2024 in the U.S. public-firm sample. The OC score ranges from 0 to 1 and measures the extent to which a job builds organization capital, based on its job description in the Revelio resume data. The AI sample includes all AI-related U.S. jobs with a valid job description. The non-AI sample consists of 15,000 randomly drawn non-AI U.S. jobs, with 1,000 jobs sampled per year from 2010 through 2024. "OC-building" denotes jobs with an OC score of at least 0.6. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table B.5. Definitions of Organization Capital Reason Categories

Panel A: Builds organization capital

Reason category	Definition
Reusable data platforms & pipelines	They build firm-specific organization capital by codifying the firm's data definitions, integration logic, governance rules, and reporting needs into reusable data platforms, pipelines, warehouses, and shared datasets that preserve operational learning and make future analytics and coordination more efficient.
Reusable AI/ML platforms & MLOps	They build firm-specific organization capital by converting the firm's accumulated model-development and deployment experience into reusable AI/ML platforms, MLOps tooling, evaluation standards, and monitoring routines that let many teams build, launch, and improve models more efficiently over time.
Firm-specific predictive & decision models	They build firm-specific organization capital by embedding firm-specific patterns, decision rules, and labeled experience into reusable predictive, ranking, recommendation, search, pricing, forecasting, and optimization models that improve future decisions as the firm learns from its own data.
Proprietary-knowledge GenAI/NLP systems	They build firm-specific organization capital by codifying proprietary documents, workflow knowledge, and judgment rules into reusable generative-AI, conversational-AI, retrieval, and NLP systems that automate recurring work and make the firm's internal knowledge easier to reuse.
Reusable software platforms & workflows	They build firm-specific organization capital by translating tacit business processes, integrations, and exception-handling know-how into reusable software platforms, APIs, enterprise applications, and automated workflows that standardize execution and reduce reliance on individual employees.
Cross-functional operating routines	They build firm-specific organization capital by institutionalizing cross-functional operating knowledge into reusable governance routines, planning cadences, delivery processes, roadmaps, documentation standards, and coordination mechanisms that make complex execution more repeatable at scale.
Reusable cloud, infrastructure & reliability operations	They build firm-specific organization capital by codifying the firm's infrastructure and reliability know-how into reusable cloud architectures, deployment pipelines, automation, monitoring, network configurations, and operating procedures that make future technology operations more scalable and dependable.
Security, compliance & risk systems	They build firm-specific organization capital by translating accumulated security, privacy, legal, compliance, and risk experience into reusable policies, controls, detection rules, audit systems, governance frameworks, and response routines that let the firm manage risk and regulated activity more efficiently over time.
Marketing & revenue-operations systems	They build firm-specific organization capital by codifying customer, market, channel, and monetization learning into reusable marketing systems, advertising models, experimentation frameworks, revenue operations tools, and commercialization playbooks that improve future targeting, launches, and sales productivity.
Customer-service & support systems	They build firm-specific organization capital by embedding customer-service and support experience into reusable workflows, self-service tools, conversational systems, escalation routines, performance metrics, and service platforms that make future service delivery more consistent and efficient.
Product & design systems	They build firm-specific organization capital by codifying product, user, and market learning into reusable product strategies, roadmaps, design systems, UX standards, lifecycle processes, and cross-functional development routines that improve the firm's future ability to create and scale offerings.
Manufacturing & process systems	They build firm-specific organization capital by converting production, quality, and plant operating experience into reusable manufacturing processes, control systems, test frameworks, automation, and cross-site routines that raise future yield, reliability, and scale efficiency.
Supply-chain & logistics systems	They build firm-specific organization capital by codifying the firm's logistics, procurement, inventory, supplier, and fulfillment learning into reusable forecasting, routing, planning, optimization, and supply-chain management systems that improve future resource allocation and operational performance.
R&D & scientific platforms	They build firm-specific organization capital by embedding scientific, clinical, experimental, and regulatory learning into reusable R&D platforms, bioinformatics pipelines, validated methods, data assets, and cross-functional development routines that make future discovery, trials, and product approvals faster and more effective.
Financial & risk decision systems	They build firm-specific organization capital by codifying financial, underwriting, fraud, portfolio, and planning knowledge into reusable data assets, models, reporting systems, and decision routines that improve future capital allocation, risk decisions, and control processes.
Hiring, training & talent development	They build firm-specific organization capital by institutionalizing the firm's people-management and skill-building know-how into reusable hiring systems, assessment methods, onboarding, training curricula, leadership practices, and talent-development routines that spread firm-specific knowledge across employees.
Alliance & partnership systems	They build firm-specific organization capital by codifying alliance, partner, acquisition, and ecosystem-management experience into reusable integration models, co-sell processes, enablement systems, governance structures, and relationship routines that improve the firm's ability to repeatedly scale through external relationships.
Knowledge organization & ontologies	They build firm-specific organization capital by organizing the firm's specialized knowledge into reusable ontologies, schemas, knowledge graphs, semantic indexes, documentation systems, and content standards that make internal information easier to integrate, search, interpret, and apply in future work.

Panel B: Does not build organization capital

Reason category	Definition
Bespoke client consulting/implementation	This work mainly delivers bespoke consulting, implementation, solution-design, or advisory services to external clients, so learning is transferred to customer-specific engagements rather than accumulating as durable firm-specific routines inside the employer.
Portable analytics & one-off models	This work mainly applies general, portable analytics, statistics, or machine-learning methods to produce project-specific models, forecasts, reports, or recommendations, so value is embodied in one-off analytical outputs and worker human capital rather than in enduring firm-specific operating knowledge.
Broadly applicable research & algorithms	This work mainly creates broadly applicable research, algorithms, prototypes, or scientific and engineering know-how that are codified and portable across employers, so the gains accrue to general technical knowledge rather than to firm-specific organizational routines.
Customer-facing products & IP	This work mainly creates customer-facing products, features, models, designs, or intellectual property whose value is embodied in marketable technology output, not in durable internal routines or process knowledge that make the employer organization itself more efficient.
Standardized/portable software work	This work mainly uses standardized software, data, or engineering tools and frameworks to produce code, systems, or technical deliverables, so the knowledge created remains largely portable across firms instead of becoming employer-specific organization capital.
Transactional sales, marketing & account management	This work mainly generates revenue through sales, marketing, partnerships, account management, or client relationship activities, so value comes from transactions and external commercial outcomes rather than from building durable internal routines or production know-how.
Individual recruiting & hiring transactions	This work mainly applies portable recruiting and labor-market matching skills to fill individual vacancies, so it produces transactional hiring outcomes rather than durable firm-specific systems, routines, or knowledge.
Standardized routine operations, support, maintenance	This work mainly executes standardized operations, support, maintenance, administrative, service, or compliance tasks within existing procedures and systems, so it sustains current output rather than creating new firm-specific knowledge that compounds into organization capital.
Project/program management (episodic)	This work mainly applies portable project, product, or program management and coordination skills to deliver discrete launches or projects, so the output is episodic execution rather than repeatable firm-specific routines that improve future operations.
Training & general education	This work mainly provides classroom-style training, rotational experience, or standalone educational projects that build general human capital, rather than firm-specific knowledge developed through learning-by-doing in the employer's own ongoing operations.
Portable legal & contracting	This work mainly applies portable legal, contracting, negotiation, or dispute-management expertise to case-specific matters, so value rests in one-off professional judgment rather than in durable firm-specific operating routines or systems.

This table reports the definitions of the reason categories tabulated in Table 1. For jobs classified as building OC, the reason category records the mechanism through which the job builds firm-specific organization capital (OC); for the remaining jobs, it records why the work does not. The LLM first induces the standardized set of categories from the job descriptions and then assigns each job to its nearest category. Panel A lists the categories of jobs classified as OC-building (OC score ≥ 0.6). Panel B lists the categories of the remaining jobs.

Table B.6. Organization Capital and AI Job Categories

Panel A: All firms

	OC-building	Non-OC-building	Jobs	Share of jobs	OC-building share
Core AI development	856	986	1,842	44.2%	46.5%
Management	741	196	937	22.5%	79.1%
Support for AI products/services	67	229	296	7.1%	22.6%
Other	662	434	1,096	26.3%	60.4%
Total	2,326	1,845	4,171	100.0%	55.8%

Panel B: Non-tech firms

	OC-building	Non-OC-building	Jobs	Share of jobs	OC-building share
Core AI development	447	452	899	46.7%	49.7%
Management	362	57	419	21.8%	86.4%
Support for AI products/services	12	34	46	2.4%	26.1%
Other	382	179	561	29.1%	68.1%
Total	1,203	722	1,925	100.0%	62.5%

Panel C: Tech firms

	OC-building	Non-OC-building	Jobs	Share of jobs	OC-building share
Core AI development	409	534	943	42.0%	43.4%
Management	379	139	518	23.1%	73.2%
Support for AI products/services	55	195	250	11.1%	22.0%
Other	280	255	535	23.8%	52.3%
Total	1,123	1,123	2,246	100.0%	50.0%

This table reports the distribution of AI-related jobs that build (OC-building) and do not build (Non-OC-building) organization capital (OC) across aggregated AI job categories for all firms (Panel A), non-tech firms (Panel B), and tech firms (NAICS 51, 54, and semiconductor manufacturing; Panel C). Jobs with an OC score of at least 0.6 are classified as OC-building. The sample consists of the classified random sample of 4,500 AI-related jobs, excluding 329 jobs without a job description. The aggregated AI-related job categories follow Table A.26. "OC-building share" is defined as the number of OC-building jobs in a given AI-related job category divided by the total number of jobs in that category.

Table B.7. Top 30 Occupations by Organization-Capital-Building Share of AI Jobs

Rank	Code	Occupation	OC-building share	OC-building jobs	Scored jobs
1	11-1021.00	General and Operations Managers	0.940	220	234
2	13-1199.04	Business Continuity Planners	0.924	281	304
3	11-3131.00	Training and Development Managers	0.922	535	580
4	11-3071.04	Supply Chain Managers	0.918	1,078	1,174
5	11-9199.02	Compliance Managers	0.917	676	737
6	11-3051.01	Quality Control Systems Managers	0.910	619	680
7	11-3021.00	Computer and Information Systems Managers	0.901	2,074	2,301
8	17-2041.00	Chemical Engineers	0.897	165	184
9	11-9179.00	Personal Service Managers, All Other	0.884	152	172
10	11-3031.00	Financial Managers	0.877	307	350
11	13-1151.00	Training and Development Specialists	0.874	841	962
12	11-9121.01	Clinical Research Coordinators	0.863	354	410
13	11-3051.00	Industrial Production Managers	0.862	119	138
14	17-2112.03	Manufacturing Engineers	0.855	271	317
15	11-3121.00	Human Resources Managers	0.850	1,320	1,553
16	13-1081.01	Logistics Engineers	0.848	139	164
17	13-1041.07	Regulatory Affairs Specialists	0.844	238	282
18	13-1141.00	Compensation, Benefits, and Job Analysis Specialists	0.843	167	198
19	13-1081.02	Logistics Analysts	0.842	336	399
20	11-3061.00	Purchasing Managers	0.839	546	651
21	13-1199.06	Online Merchants	0.833	275	330
22	13-2031.00	Budget Analysts	0.833	90	108
23	13-2054.00	Financial Risk Specialists	0.833	1,273	1,529
24	17-2112.00	Industrial Engineers	0.830	420	506
25	25-9031.00	Instructional Coordinators	0.820	105	128
26	15-1212.00	Information Security Analysts	0.813	693	852
27	17-2112.02	Validation Engineers	0.805	231	287
28	13-2099.04	Fraud Examiners, Investigators and Analysts	0.804	197	245
29	43-3061.00	Procurement Clerks	0.800	140	175
30	15-1299.05	Information Security Engineers	0.794	1,991	2,506

This table reports the thirty O*NET occupations with the highest share of AI-related jobs classified as OC-building. The sample consists of AI-related U.S. jobs with valid descriptions in 2024 in the U.S. public-firm sample. The sample is restricted to firm-years with at least 50 U.S. workers in Revelio and at least 50 employees in Compustat. Occupations with fewer than 100 AI-related jobs are excluded. "OC-building share" is defined as the number of scored AI-related jobs in an occupation classified as OC-building divided by the total number of scored AI-related jobs in that occupation. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table B.8. Top 30 Occupations by Number of Organization-Capital-Building AI Jobs

Rank	Code	Occupation	OC-building share	OC-building jobs	Scored jobs
1	15-1252.00	Software Developers	0.556	21,667	38,953
2	15-2051.00	Data Scientists	0.573	20,854	36,389
3	11-2021.00	Marketing Managers	0.710	11,878	16,737
4	15-1299.09	Information Technology Project Managers	0.777	11,750	15,127
5	15-2051.01	Business Intelligence Analysts	0.720	10,001	13,882
6	15-1243.00	Database Architects	0.752	8,299	11,036
7	15-1299.08	Computer Systems Engineers/Architects	0.662	7,760	11,724
8	15-1221.00	Computer and Information Research Scientists	0.482	7,655	15,888
9	17-2199.06	Microsystems Engineers	0.690	5,461	7,914
10	15-1255.00	Web and Digital Interface Designers	0.636	3,658	5,753
11	11-1011.00	Chief Executives	0.691	2,263	3,276
12	17-2199.00	Engineers, All Other	0.710	2,088	2,941
13	11-3021.00	Computer and Information Systems Managers	0.901	2,074	2,301
14	15-1299.05	Information Security Engineers	0.794	1,991	2,506
15	13-1071.00	Human Resources Specialists	0.467	1,883	4,035
16	13-1199.00	Business Operations Specialists, All Other	0.788	1,517	1,925
17	13-1161.01	Search Marketing Strategists	0.765	1,432	1,872
18	41-9031.00	Sales Engineers	0.317	1,407	4,436
19	11-3031.01	Treasurers and Controllers	0.719	1,390	1,933
20	23-1011.00	Lawyers	0.642	1,387	2,161
21	15-1253.00	Software Quality Assurance Analysts and Testers	0.731	1,321	1,807
22	11-3121.00	Human Resources Managers	0.850	1,320	1,553
23	13-2054.00	Financial Risk Specialists	0.833	1,273	1,529
24	15-1243.01	Data Warehousing Specialists	0.752	1,257	1,671
25	13-1161.00	Market Research Analysts and Marketing Specialists	0.623	1,226	1,969
26	11-2022.00	Sales Managers	0.465	1,187	2,552
27	17-2072.01	Radio Frequency Identification Device Specialists	0.621	1,153	1,856
28	15-1299.07	Blockchain Engineers	0.776	1,147	1,478
29	15-1254.00	Web Developers	0.515	1,112	2,158
30	27-1029.00	Designers, All Other	0.481	1,091	2,267

This table reports the thirty O*NET occupations with the largest number of AI-related jobs that build organization capital (OC-building). The sample consists of AI-related U.S. jobs with valid job descriptions in 2024 in the U.S. public-firm sample. The sample is restricted to firm-years with at least 50 U.S. workers in Revelio and at least 50 employees in Compustat. Occupations with fewer than 100 AI jobs are excluded. "OC-building share" is defined as the number of scored AI jobs in an occupation classified as OC-building divided by the total number of scored AI jobs in that occupation. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table B.9. Examples of AI Jobs by Organization-Capital Classification

Job title	Job description	Reason category
Panel A: Organization-capital-building AI jobs		
Director and Principal Data Scientist, Next Gen Services (NGS), Global Business Services	CTO for AI and ML-driven "Exponential Value" innovations for corporate business services. Corporate Knowledge Management using Generative AI: Insight engine for explicit knowledge using large language models. Capture and insight building from tacit knowledge residing with Experts (SpeechBrain). Real-time Digital Twins of business processes: Supply chain logistics digital twins and risk predictions. Customer order management digital twin to minimize on-shelf out-of-stocks, smoothing truck deliveries, and optimize plant production processes. Continuous auditing digital twin, failure and incident prediction, and prescription. Procurement Guidance Engine: Price prediction for raw materials and commodities. Supplier Risk Assessment § Financial forecasting. AI@Scale: Model Governance and Monitoring with an aim to create a digital twin of MLOps.	Proprietary-knowledge GenAI/NLP systems
Vice President	Architected a GenAI-powered agent to analyze call transcripts and flag unaddressed items for mortgage Operations; transformed from a hackathon concept into a funded enterprise initiative. Designed and deployed an LLM-based mortgage underwriting assistant to standardize underwriting memos, projected to save over 10,000 hours annually. Built a Prompt Engineering Platform with integrated scoring, approval workflows, and audit tracking-reducing prompt creation overhead by 40%. Designed and delivered a real-time OLTP reporting platform using Kafka and Power BI to enable low-latency insights and increasing mortgage operational visibility. Established and scaled a data lineage practice within Banking Products, increasing adoption 5x and cutting Collibra onboarding costs by 40%. Serve as the Data Domain Architect for Wealth Banking PAG, reviewing and guiding data-related architecture and implementation across lending and banking systems. Developed data design patterns for tooling selection, lineage accessibility, data quality management, purging standards, and audit readiness. Led enterprise lending data strategy and defined a 3-year vision across governance, entitlements, tooling, and holistic data management. Driving KYC banking data integration across 16+ systems, enabling a strategic shift to client-level screening and streamlined periodic reviews.	Financial & risk decision systems

Job title	Job description	Reason category
Director of Trading Surveillance and Data Strategy	Designed and built a scalable statistical surveillance platform with profiling, Complex Event Processing and real time capabilities to address surveillance challenges that are non-deterministic in nature. Conceptualized and initiated a project to use machine learning and Big Data to connect data points across trading and E-Communications. Appointed as the Compliance Data Officer for Institutional Compliance globally. Responsible for all data strategy including data acquisition and on-going data quality monitoring.	Security, compliance & risk systems
Program Director, Honeywell Competency Management	Selected by Honeywell executive management to lead the development and commercialization of Honeywell's Competency Management practice focused on building and maintaining employee skills while transitioning educational methods to meet the needs of the next generation of workers. Leveraged IoT technologies including AR/VR, AI, wearable and other smart technologies to capture and transfer knowledge and skills. Managed the development of advanced analytics designed to assess employee skills as they perform their job on an ongoing basis. Created the contracts, business processes, KPI Scorecards and metrics required to support the delivery of a formal competency program.	Hiring, training & talent development
Computer Vision AI/ML Engineer	As a core member of the global Automatic Visual Inspection (AVI) project, provide image analysis solutions, including deep learning and rule-based algorithms. Lead and standardize the MLOps ecosystem, tools and processes to support the development and deployment of AI/ML algorithms for AVI solutions. Build communication platforms and lead forums to promote knowledge sharing and increase the visibility of the AVI project, while achieving technology promotion and application in more than 80 manufacturing bases around the world.	Firm-specific predictive & decision models
UX Director, AI	Lead AI design team + strategy across HubSpot's product suite. Define principles, frameworks, and interaction models for integrating AI into CRM and workflow tools. Grew AI UX discipline 0->20 people in 2 years.	Product & design systems

Job title	Job description	Reason category
	Shaped the design of Breeze, an AI-powered product line, including agents and assistants for integrated SMB AI use-cases Developed approach to conversational interfaces, AI collaboration and AI UX patterns Helped evolve HubSpot's design system to support AI-native features Collaborated across product, engineering, and marketing to align AI capabilities with strategic goals Driver of company wide adoption for AI product design practices, principles and tool adoption	
VP, Data, AI, Robotics Platforms	Designed the data platform strategy, enabling data-driven decisions across all organizational levels. Developed and launched an AI platform integrating LLMs (e.g., OpenAI) with proprietary data, both structured and unstructured, streamlining operations and uncovering actionable insights. Spearheaded the design and implementation of advanced data science applications, including an operational control tower, significantly transforming supply chain and reducing vaccine distribution costs and improved service level. Led the robotics platform to enhance adoption and innovation in automation, with a primary focus on personalized cancer vaccine manufacturing and distribution.	Reusable data platforms & pipelines
Product Management, AV MLOps	I lead the Product Management team within MLOps for Autonomous Vehicles within NVIDIA. We're enabling data flywheels for machine learning engineers across perception, planning/control, GT generation, and more!	Reusable AI/ML platforms & MLOps
GenAI Engineer - D365 Sales AI	Delivered an enterprise GenAI Research & Policy Copilot (RAG + agent workflows) used by 600+ analysts/associates, handling 12K+ queries/week, reducing document search time by 35-45% across Wealth/Risk/Compliance teams. Built an end-to-end RAG ingestion pipeline (Python, Spark/Databricks) to process 2.5M+ documents (PDFs, policies, research notes), achieving 99.5% indexing success, cutting refresh time from 24 hrs 2 hrs with incremental updates. Implemented hybrid retrieval (BM25 + embeddings) with reranking, metadata filters, and query rewriting; improved Top-3 retrieval precision from 62% 82% (+20 pts) on a 1,200-question evaluation set. Reduced hallucinations by 40% using citation-first prompting, grounding checks, and answer validation; achieved 95%+ citation coverage and <2% "no-source" responses in production telemetry.	Proprietary-knowledge GenAI/NLP systems

Job title	Job description	Reason category
	Developed agentic tool-calling workflows (LangChain/LangGraph) integrating 25+ internal APIs (portfolio/risk/reference data), cutting manual analysis steps by 50% while enforcing RBAC and generating 100% audit logs for regulated use cases. Implemented security & compliance guardrails: PII/PCI redaction (99% recall), prompt-injection defenses with 98% block rate on red-team test suites, and policy-based routing for restricted content aligned with SEC/FINRA controls. Established LLMOps (MLflow/Prompt versioning, automated evals, regression tests, safety checks) that reduced release cycles from 2 weeks 3 days and increased automated coverage to 90%+ across prompts/retrieval/config. Optimized inference performance via semantic caching (Redis), response streaming, dynamic context sizing, and model routing; improved p95 latency from 6.2s 2.1s and lowered compute cost per query by 45%. Deployed containerized GenAI services on AKS/Kubernetes using CI/CD (GitHub Actions/Azure DevOps) + IaC (Terraform), achieving 99.9% uptime, supporting 3 traffic spikes, and reducing MTTR by 30% using OpenTelemetry-based tracing.	
Software Engineer - Speech AI, Evals	Led evaluation and fine-tuning of transformer based in-house multilingual Automatic Speech Recognition (ASR) models, including Whisper and Wav2Vec, enabling them to process over 34 million+ daily requests with high accuracy. Prototyped a large-scale annotation, ranking, and auto-evaluation pipeline using multiple ASR engines + LLM-as-judge, enabling scalable and reliable ground-truth generation. Designed a synthetic query generation system (LLM + RAG) to produce context-aware, policy-violating test cases, stress-testing conversational AI guardrails and validating model robustness. Also introduced a probability drift metric for early detection of ASR performance degradation, reducing customer-impacting issues. Engineered "Sherlock," an in-house model evaluation framework built with Java 17, significantly enhancing ASR model assessments. My contributions also include deploying ASR models in production and QA environments that mirror production, reducing pre-deployment issues by 40%, and building an automated testing pipeline that processes 10,000+ production audio files daily, ensuring system reliability under production-level loads.	Knowledge organization & ontologies

Job title	Job description	Reason category
Panel B: Non-organization-capital-building AI jobs		
Managing Consultant	Management Consultant leveraging Big Data and AI to solve business problems at clients like PepsiCo, Amgen, Genentech and SuperValu	Bespoke client consulting/implementation
Equity Research Analyst (buy-side)	Built and maintained models on all portfolio holdings and companies on watchlist (∼200 companies) Invested in Industrial, Energy, and Consumer sectors with a focus on DCF, share buybacks, and operating margin mean reversion Incorporated technical and market sentiment analysis to identify optimal trading entry and exit points Ran a mock L/S equity portfolio to track my stock recommendations which outperformed its index by 34.7% over 5 months	Portable analytics & one-off models
Data Scientist	Build and optimize machine learning models for analyzing massive amounts of data and providing business insights. Perform data processing, feature engineering, and model evaluation using Python and R. Collaborate on the development of end-to-end data science solutions to provide personalized recommendations to clients.	Portable analytics & one-off models
Software Engineer	During my time at Interphase, I worked on a new product (Penveu) based around the use of computer vision and sensor fusion in a handheld form factor. Designed and implemented a method of hiding and retrieving imperceptible patterns in images displayed on computer monitors and projection displays Studied and provided design parameters for component specifications, including human design considerations	Customer-facing products & IP
Software Engineer	Participated in the development of Oracle cloud computing platform, focusing on the implementation of big data analysis and machine learning functions. Developed distributed data processing systems using Java and Python, optimized customers' business processes and increased data processing speed. Designed and deployed multiple data analysis tools to help corporate customers better understand and analyze their business data. Participated in cross-departmental projects to improve product scalability and support the global deployment of cloud computing platforms.	Standardized/portable software work
Senior Staff Software Engineer	Design, implementation and optimization of hardware VP9 encoders / decoders. Chief in inventing and engineering world class VP9 4K60-capable (single-core, 600MHz) hardware encoder, some of the fancier features being	Customer-facing products & IP

Job title	Job description	Reason category
	• practically unlimited motion search window over 3 ref frames. • RD-optimal mode selection. • trellis quantization. • temporal denoising.	
Distinguished Research Scientist	Advanced robotics research in machine learning, motion generation, control, and integrated systems for real-world dextrous manipulation. At NVIDIA's robotics lab in Seattle we're driving core algorithmic technologies toward the democratization of advanced robotics systems to unlock new applications across manufacturing, logistics, healthcare, service, and homes. With some of the world's foremost experts in bleeding edge algorithms for robotics, we're driving toward a unique and deep understanding of robot intelligence and pushing this technology toward unprecedented performance.	Broadly applicable research & algorithms
Research Scientist	Responsible for developing advanced and reliable industrial software for automation applications such as smart manufacturing, industrial robots and cyber physical systems. Technical expertise includes automated reasoning, automatic control and artificial intelligence.	Broadly applicable research & algorithms
AI Engineer	Responsible for designing and developing highly scalable artificial intelligence and machine learning applications for industrial, energy and financial applications. Utilize technologies such as deep learning, natural language processing and image recognition to solve complex business problems faced by customers.	Portable analytics & one-off models
Account Executive, Small Business	Sold Salesforce's multi cloud and AI products to new and existing customer accounts 262% FY25 ACV attainment ($1.1 M closed ACV), 130% FY25 Multiyear attainment, 92% FY25 New Logo attainment Recognized as South MVP for FY25 for exceptional performance; ranked Top 5 in Small Business (Americas) and earned the South 6 Achievement Award Managed accounts across the U.S. in key sectors including professional services, technology, manufacturing, engineering, and energy	Transactional sales, marketing & account management

This table shows examples of AI jobs classified as OC-building (Panel A) and non-OC-building (Panel B), together with each worker's job title, job description, and reason category. *GPT-5.4* reads each job's title and description, assigns an organization capital score between zero and one, and classifies the job as OC-building if its score is at least 0.6. The reason category captures the mechanism behind the classification. See Appendix Table B.5 for definitions of each category.

Table B.10. Examples of Non-AI Jobs by Organization-Capital Classification

Job title	Job description	Reason category
Panel A: Organization-capital-building non-AI jobs		
Restaurant Guest Service Manager of Disney's Flying Fish Cafe	As RGSM at the Flying Fish, I led and supported a cast of approximately 50 servers, seaters and busers to ensure that Disney standards and basic was followed at all times, I educated cast members on food safety and health procedures and familiarized them with how best to ensure these standards are met in our facility. I dedicated a significant amount of time and energy to being visible on the floor, interacting with guests and demonstrating our dedication to providing them with an unforgettable dining experience. Accomplishments include re-organizing our wine storeroom in compliance with the updated receiving standards. In this position, I worked to further the development of our seasoned cast members through pre-meal coaching and one-to-one meetings.	Hiring, training & talent development
EVP/Head of the Corporate Risk Program Office	Head of the Corporate Risk Program Office responsible for management strategy and execution across Wells Fargo's various functional areas, platforms, and risk disciplines. She works to strengthen the risk management function across the company and ensure that the risk perspective is integrated with major initiatives, including those driving improvements in the customer experience.	Security, compliance & risk systems
Senior Software Engineer	I began working in software configuration management and have moved to performance simulation where I have been working in build management. I am the current build management team lead, leading a team of 6 of us. I am responsible for managing our regression testing and test scripts, planning out the direction of our testing and capabilities for multiple programs, as well as defining the overall process and development life-cycle for the program.	Product & design systems
Talent Portfolio Delivery Planner	Designing and planning delivery for enterprise talent experiences Create and maintaining global delivery standards Fostering delivery innovation by sensing and incorporating industry best practices. Partnering with regional teams to plan and execute enterprise talent experiences Being the single partner contact for the Talent Management Organization talent segment and design partners.	Hiring, training & talent development

Job title	Job description	Reason category
Site Quality Manager	Created a centralized Quality System that allowed continued operation of ISO 17025, NADCAP, ISO 9001, and IATF Certifications under a 75% headcount reduction as the only person with a quality background on site. Developed and sourced material coupon manufacturing and inspection from a 3rd party facility in 5 weeks with no on-site support under the immediate loss of the current in house facility.	Manufacturing & process systems
Operations Manager II	Controls specific day-to-day operations which include training, recognition, coaching, performance evaluations, safety training, and staff productivity Manage a team of 50+ Manage 120 truck and delivery routes Administer and adhere to company safety strategies and programs Develop and support new operational strategies to improve overall success and safety	Hiring, training & talent development
Fulfillment Supervisor	Led outbound sortation and shipping in a 1M+ sq. ft. Amazon FC with 250+ associates and 4 direct shift leads. Improved productivity and reduced overtime by 18% through labor optimization and real- time monitoring. Implemented workflow changes based on A/B testing and Six Sigma methodology. Built high-performing teams by creating mentorship and peer-coaching programs. Partnered with HR and Loss Prevention to address performance, attendance, and integrity cases. Oversaw inventory reconciliation efforts and resolved discrepancies in coordination with ICQA team.	Supply-chain & logistics systems
Sales Facilitator	New Hire facilitation for both Residential Voice and Residential Chat. Facilitate learning to develop participants' skill sets. This includes classroom instruction, team meetings, and on-line Learning debriefs. Perform all administrative duties including class and lab preparation, administration, and participant record maintenance. Coach learners on specific skills within a role, process or system. Develop new hire curriculum for effectiveness and appropriate content.	Hiring, training & talent development

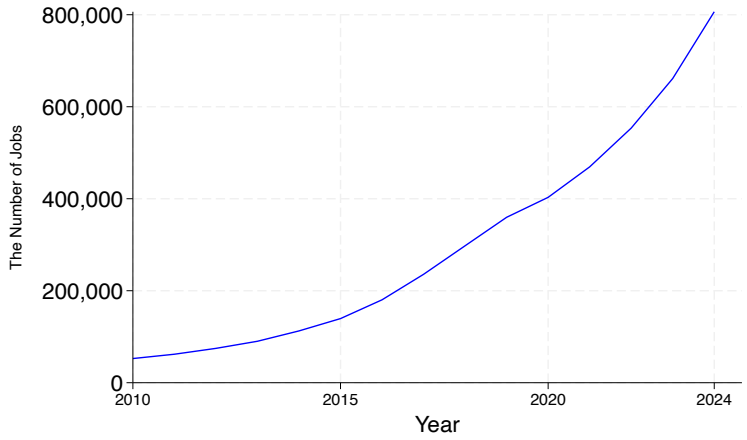
Job title	Job description	Reason category
Customer Experience Manager	• Collaborate with Cross-functional teams to align customer experience with business objectives. • Ensured comprehensive trainings are delivered to internal teams for newly launching processes. • Collaborate with Care and Sales to streamline processes and optimize workflow within Government Education Programs. • Conduct Steering Meetings to collaborate and align on program goals. • Conducted a thorough analysis of existing processes and workflows to identify obstructions and areas of optimization. • Work as a SME for Government processes to launch new initiatives.	Customer-service & support systems
Panel B: Non-organization-capital-building non-AI jobs		
National Account Executive - Field Sales	Online advertising sales. Prospect for and grow national advertising partners. Assist in driving sales from the largest real estate audience in the US (37 million unique users/mo). My specific areas of focus have been homes builders, real estate brokerages and auction houses.	Transactional sales, marketing & account management
Copy and Print Associate	Worked with customers to assess the needs of their particular project. Experience operating photocopiers, large format printers, book binders, and cutting machines. Performed quality checks on customer orders and worked with others to complete orders. Used programs such as Adobe Acrobat, Word, Publisher, Excel, Cute PDF, and Photoshop. Maintained a work log that helped to prioritize workflow.	Standardized routine operations, support, maintenance
Mortgage Resolution Analyst	Analyze attorney firms billing invoices for accuracy and confirm invoices are received within the allowable timeframe. Research denied invoices for payment resolution. Review and verify accuracy of invoices and supporting documentation prior to processing. Submit invoices to Account Payables for payment. Ability to process a high volume of invoices per month.	Standardized routine operations, support, maintenance
Licensed Health Insurance Broker and District Manager	Licensed Health Insurance Broker (June 2014-Present) Delivered high impact educational presentations regarding law changes, The Affordable Care Act, product overviews, and how clients can streamline back office administration Performed quote reviews with client to choose appropriate plan options to satisfy and retain employees Responsible for leading and coordinating all Health Insurance enrollments for new clients	Transactional sales, marketing & account management

Job title	Job description	Reason category
Trailer Mechanic	Attend to all aspects of trailer maintenance and repair according to company procedure. Restore, lubricate, and adjust hydraulic drum & shoe brake systems. Diagnose and repair electrical lighting issues. Mend damages in trailer body and frame using an angle grinder and MIG welder. Safely use a variety of hand, power, and pneumatic tools while utilizing proper PPE.	Standardized routine operations, support, maintenance
Web Content Writer	I create compelling content and communications strategy for technology, science, retail, health and legal companies ranging from start-ups to Fortune 500 companies. I write and edit blog posts, white papers, web copy, thought leadership pieces, trend articles, press releases, brand copy, social media content, newsletters, internal publications, online questionnaires and video scripts.	Customer-facing products & IP
Internet/VoIP Support Technician	Resolved customer connectivity issues related to internet and wireless network configurations, ensuring stable and reliable service. Assisted in porting phone numbers from other carriers, streamlining transitions for VoIP users. Collaborated with internal teams to deliver seamless customer experiences and address complex system issues.	Standardized routine operations, support, maintenance
Customer Service Representative	During my time at Sykes, I worked for the PNC Customer Care Center. I would service an average of 30-40 customers with a range of issues that they had with their bank account. Some of the concerns include: balance discrepancies, online banking troubleshooting, identity theft prevention, dispute management & other responsibilities.	Standardized routine operations, support, maintenance
Cashier	Receiving payments and issuing receipts Keeping track of all cash and credit transactions Communicating effectively with customers Using point-of-sale technologies, such as cash register, computer, barcode scanners, etc.	Standardized routine operations, support, maintenance

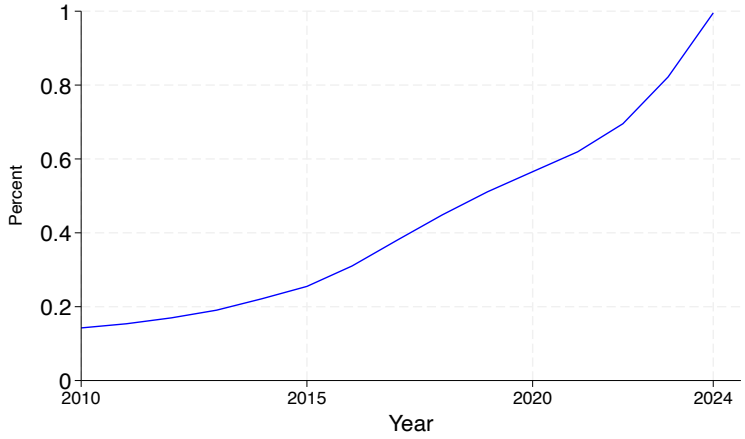
This table shows examples of non-AI jobs that build organization capital (OC-building; Panel A) and do not build organization capital (Non-OC-building; Panel B), together with each worker's job title, job description, and reason category. The non-AI sample is a representative random sample of 15,000 non-AI jobs drawn from the same firm-years as the AI-related jobs. *GPT-5.4* reads each job's title and description, assigns an organization capital score between zero and one, and classifies the job as OC-building if its score is at least 0.6. The reason category identifies the mechanism through which a job builds OC or why it does not build OC. Appendix Table B.5 provides definitions of each reason category.

C. Robustness and Additional Results

Figure C.1. Time Series of AI Jobs (Non-Public Firms)



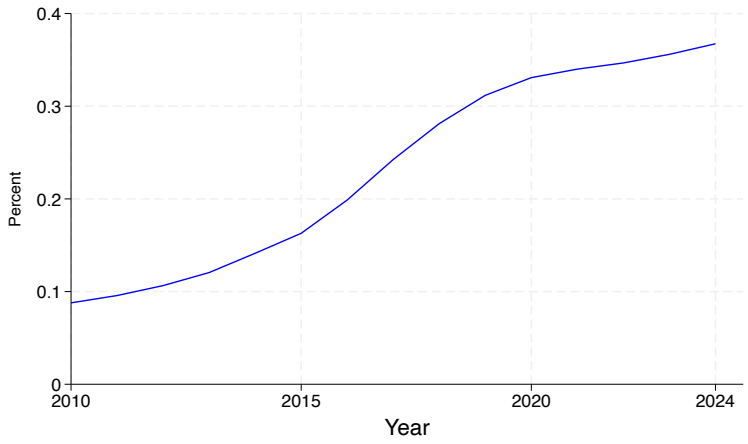
(a) Number of AI Jobs



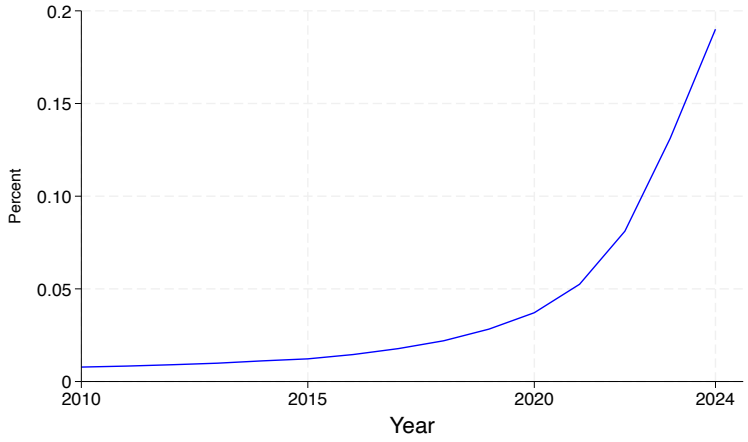
(b) Share of AI Jobs

This figure plots annual AI-related U.S. jobs among employers that are not U.S. public firms in Revelio from 2010 to 2024. These employers are mostly privately held firms, government, and non-profit sector entities. Panel (a) reports the total number of AI-related U.S. jobs. Panel (b) reports the aggregate AI job share, defined as AI-related U.S. jobs divided by total U.S. jobs in the same year, expressed in percentages.

Figure C.2. Share of AI Jobs: Wave I vs. Wave II AI Technologies (Non-Public Firms)



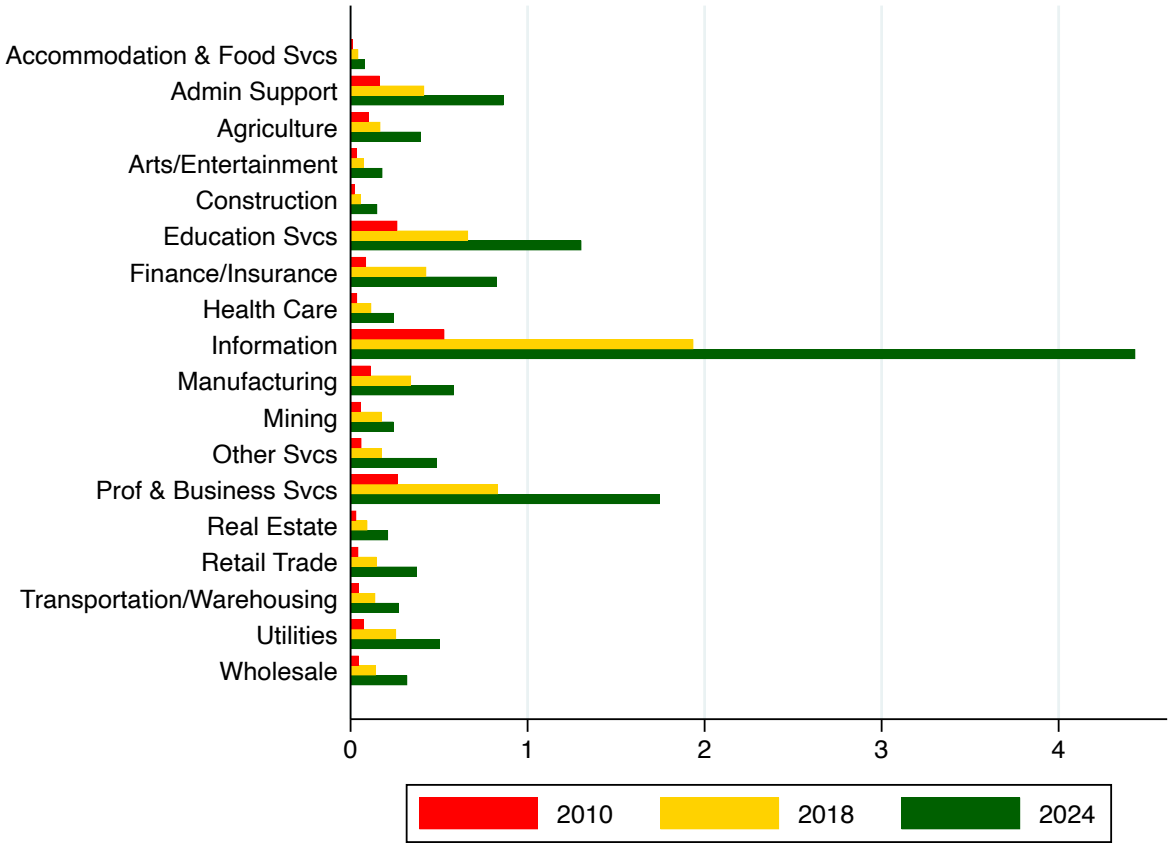
(a) Wave I: ML, NLP, and CV



(b) Wave II: Gen AI and Agentic AI

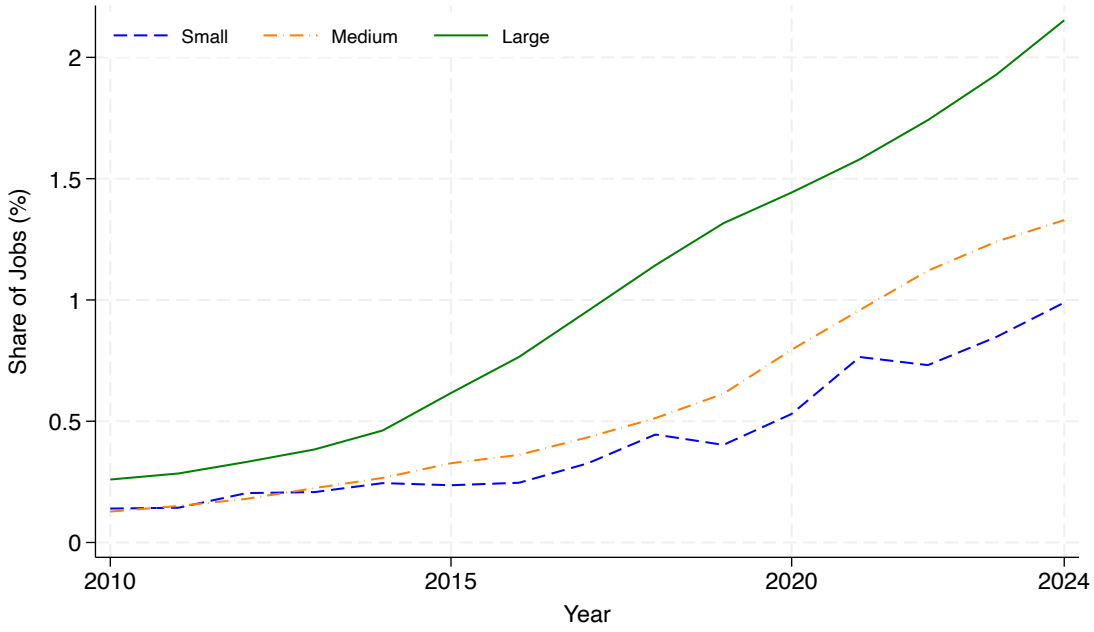
This figure plots the share of AI-related jobs among employers that are not U.S. public firms in Revelio from 2010 to 2024, separately for Wave I and Wave II AI technologies. These employers are mostly privately held firms, government, and non-profit sector entities. Panel (a) reports the share of "Wave I" AI jobs, defined as jobs with job titles or descriptions containing at least one of the keywords from machine learning (ML), natural language processing (NLP), or computer vision (CV) technology subcategories. Panel (b) reports the share of "Wave II" AI jobs, defined as jobs with job titles or descriptions containing at least one of the keywords from generative AI (Gen AI) or agentic AI technology subcategories. Each share equals U.S. AI jobs in the indicated wave divided by all U.S. jobs in the same year, expressed in percentages. Wave I and Wave II are not mutually exclusive.

Figure C.3. Share of AI Jobs by Industry Sector (Non-Public Firms)



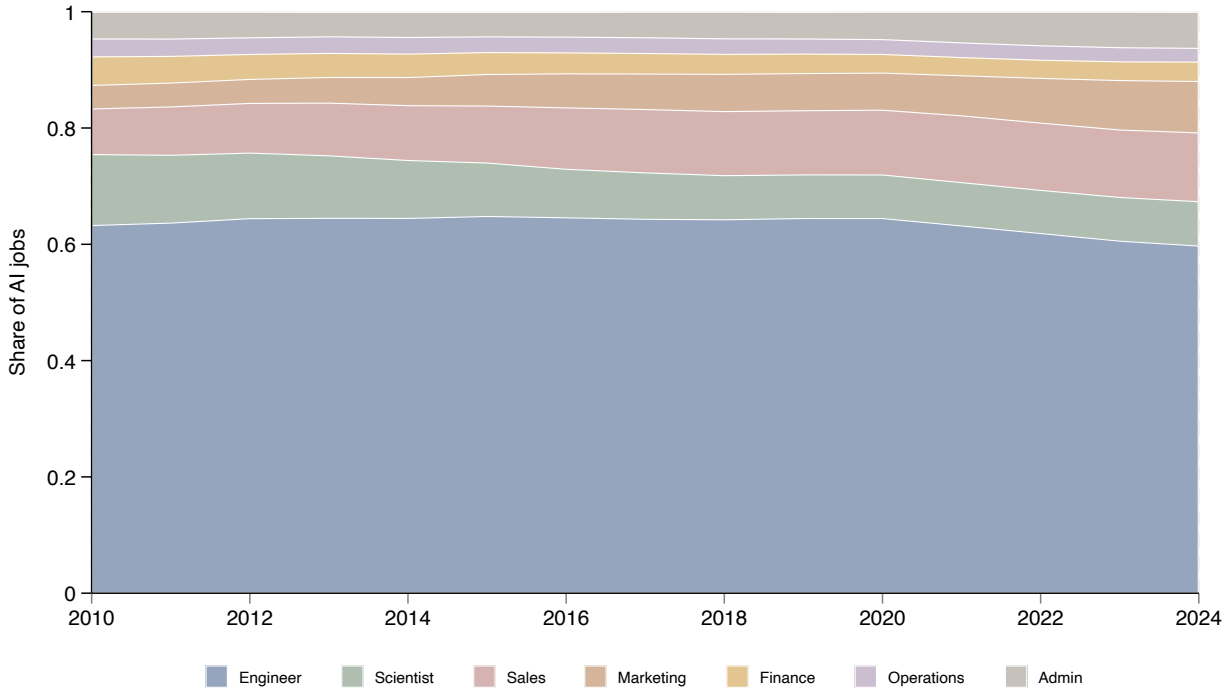
This figure presents the share of AI-related U.S. jobs at the industry level, based on the sample of jobs among employers that are not U.S. public firms in Revelio. These employers are mostly privately held firms, government, and non-profit sector entities. For each sector (based on 2-digit NAICS industry codes) and year (2010, 2018, and 2024), the figure reports the ratio of the total number of AI-related U.S. jobs in that sector divided by the corresponding total number of U.S. jobs in that sector, expressed in percentages.

Figure C.4. Share of AI Jobs by Firm Size Tercile



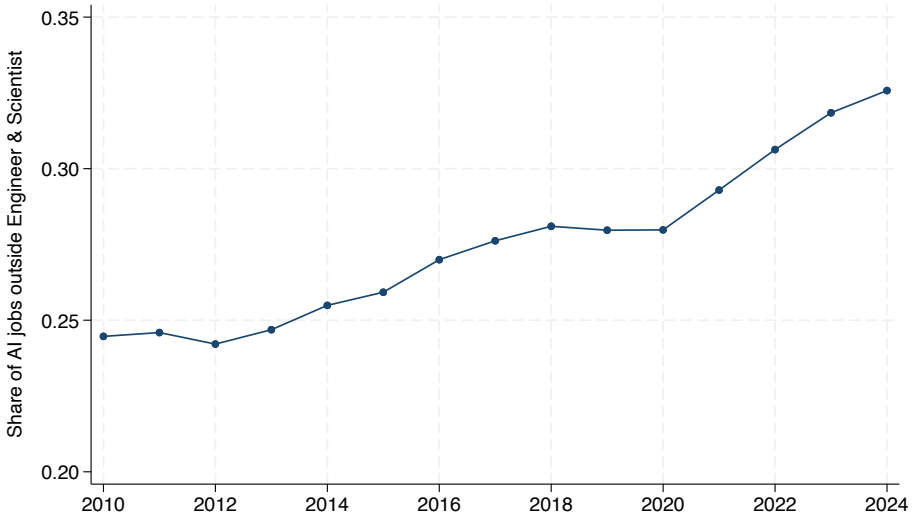
This figure plots the aggregate share of AI-related U.S. jobs by firm-size tercile from 2010 to 2024 in the U.S. public-firm sample. Each year, firms are sorted into terciles based on total global Revelio job counts. Small, Medium, and Large denote the bottom, middle, and top terciles, respectively. Within each tercile-year cell, AI share is defined as AI-related U.S. jobs divided by total U.S. jobs across all firms in that cell, expressed in percentages. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure C.5. Share of AI Jobs by Business Function



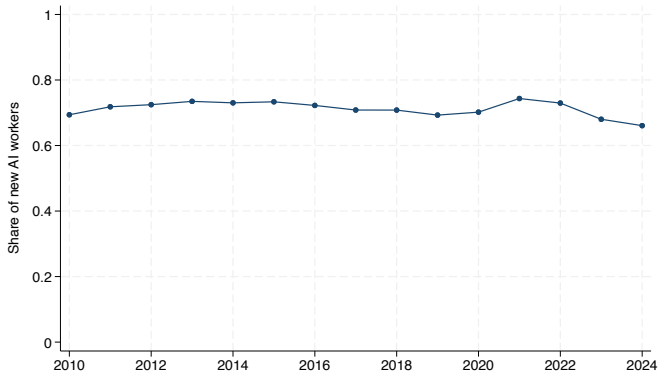
This figure plots the composition of AI-related U.S. jobs in the U.S. public-firm sample across seven broad job functions defined by Revelio from 2010 to 2024. Job functions are as reported by Revelio, which classifies each position into one of seven broad business functions—Engineer, Scientist, Sales, Marketing, Finance, Operations, and Admin—based on its job role. Within each year, each function's share equals that function's AI-related U.S. jobs divided by all AI-related U.S. jobs in the sample; the stacked areas sum to 1. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Figure C.6. Share of AI Jobs Outside Engineer and Scientist Functions

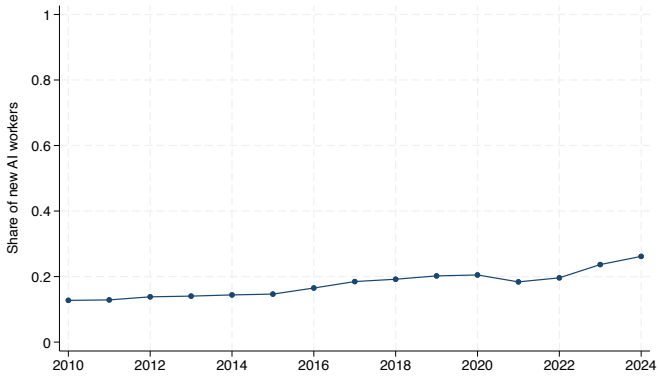


This figure plots the share of AI-related U.S. jobs from 2010 to 2024 in the U.S. public-firm sample whose job function is outside the Engineer and Scientist functions, defined as the ratio of the number of AI-related jobs outside Engineer and Scientist divided by total AI-related job count. Job functions are as reported by Revelio, which classifies each position into one of seven broad business functions—Engineer, Scientist, Sales, Marketing, Finance, Operations, and Admin—based on its job role. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

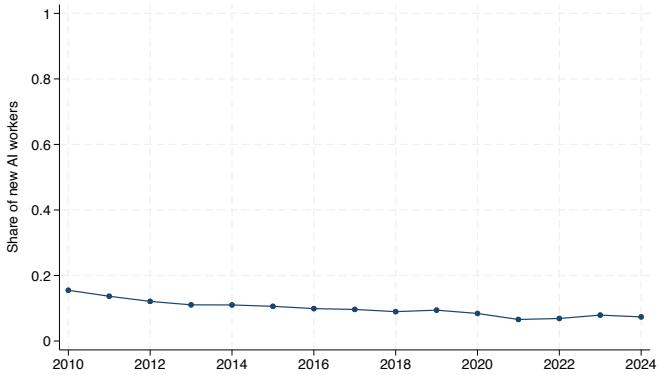
Figure C.7. Origins of New AI Workers



(a) Retrained from U.S. Jobs



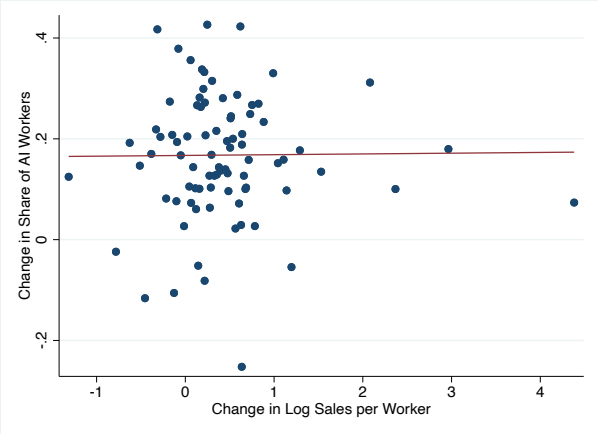
(b) Recent U.S. Graduates



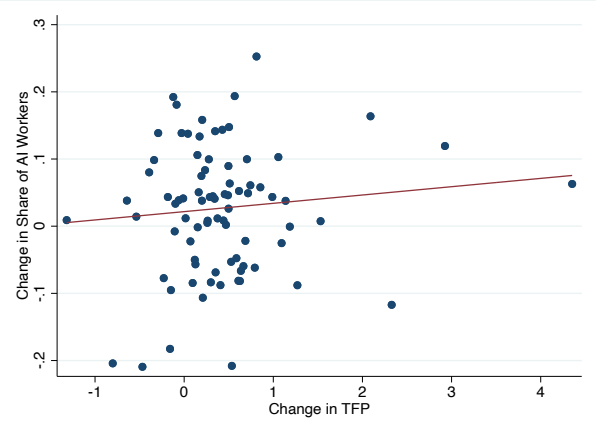
(c) Other

This figure classifies workers who begin their first AI-related job at a U.S. public firm in each year 2010–2024 by their history prior to that job, using their full Revelio employment records (including positions at non-public employers) and education histories. Each panel plots one origin group's share of that year's new AI workers. Panel (a): workers with prior U.S. non-AI work experience. Panel (b): workers whose most recent degree, from a U.S. university, was completed within one year of starting the AI job. Panel (c): workers whose most recent degree, from a non-U.S. university, was completed within one year of starting the AI job, and workers with neither prior U.S. employment nor a U.S. degree before the AI job. The residual group of workers who do not belong to any of the above three categories (e.g., workers with a U.S. degree completed more than one year before the AI job but no recorded U.S. employment prior to it) is about 1% throughout 2010–2024. Firm-years are restricted to those with a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

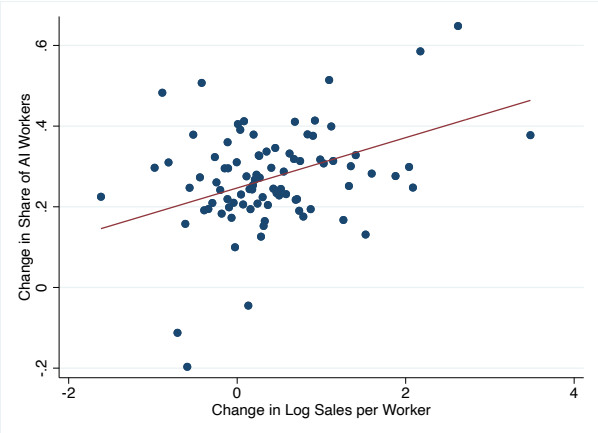
Figure C.8. Relationship Between AI Investments and Productivity Growth



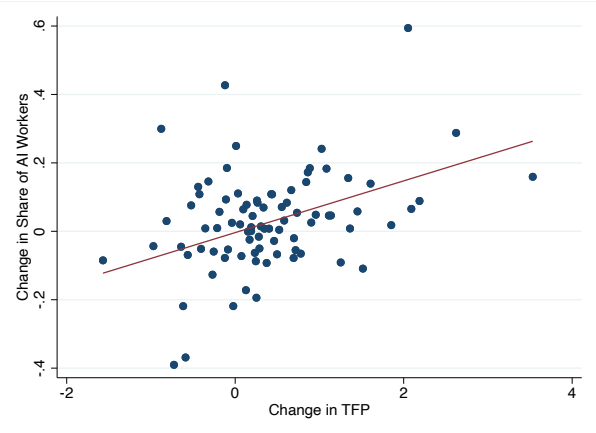
(a) Log sales per worker, 2010–2018



(b) TFP, 2010–2018



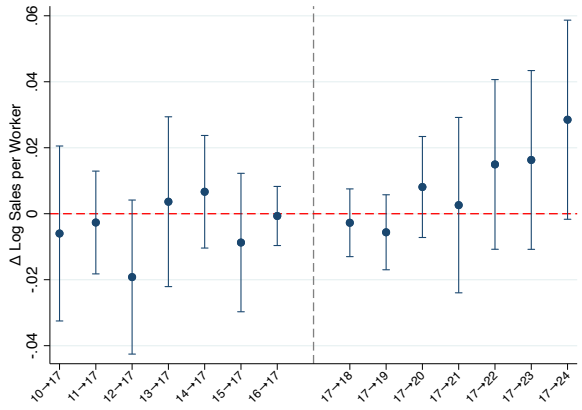
(c) Log sales per worker, 2018–2024



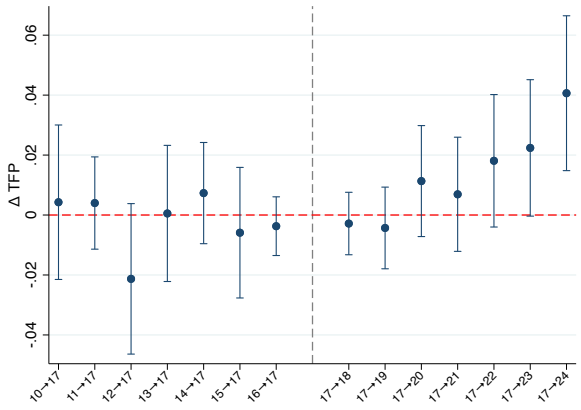
(d) TFP, 2018–2024

This figure is a binned scatterplot of the change in log sales per worker or revenue TFP against the change in the share of AI jobs for U.S. public firms (in non-tech sectors). All changes are measured from 2010 to 2018 in Panels (a) and (b) and from 2018 to 2024 in Panels (c) and (d). Each dot represents roughly the same number of firms, and the solid line is the fitted regression line. Observations are weighted by the number of jobs in Revelio at the start of the period. All specifications control for 2-digit NAICS industry sector fixed effects and baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers).

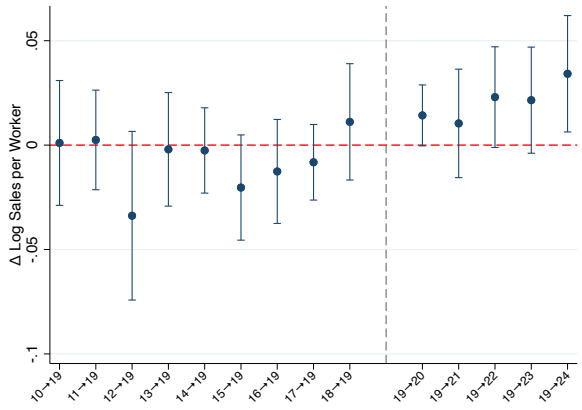
Figure C.9. AI Investments and Productivity Changes Over Time: Alternative Base Years



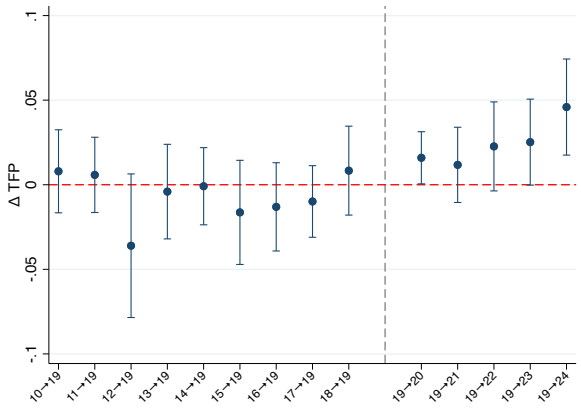
(a) Log Sales per Worker, Base Year 2017



(b) TFP, Base Year 2017



(c) Log Sales per Worker, Base Year 2019



(d) TFP, Base Year 2019

This figure plots the association between AI investment intensity and productivity growth over different windows in U.S. public firms (in non-tech sectors), using two alternative base years. In Panels (a) and (b), the independent variable is the standardized change in a firm's share of AI jobs from 2017 to 2024, $\Delta\text{Share AI Jobs}_i^{17\to24}$; in Panels (c) and (d), it is the corresponding change from 2019 to 2024, $\Delta\text{Share AI Jobs}_i^{19\to24}$. Within each pair of panels, the independent variable is the same across all regressions. Each point plots the coefficient $\hat{\beta}$ from a separate cross-sectional regression of $\Delta Y_i^{T\to T'}$ on the change in the AI job share, controlling for 2-digit NAICS sector fixed effects and weighted by the number of jobs in Revelio in the base year. Panels (a) and (c) report results for log sales per worker; Panels (b) and (d) report results for revenue total factor productivity (TFP). Let B denote the base year (2017 in the top row, 2019 in the bottom row). To the left of the dashed line, the dependent variable is $\Delta Y_i^{T\to B}$ for $T = 2010, \ldots, B-1$. The first dot measures how changes in firm AI investments from B to 2024 relate to the ex-ante change in productivity from 2010 to B; each successive dot shortens the window over which ex-ante productivity is measured by one year, with the last dot on the left using the change from $B-1$ to B. To the right of the dashed line, the dependent variable is $\Delta Y_i^{B\to T}$ for $T = B+1, \ldots, 2024$: the first dot after the dashed line uses the change in productivity from B to $B+1$, and the last dot uses the change from B to 2024. The vertical bars indicate 95% confidence intervals. Standard errors are clustered at the 5-digit NAICS industry level.

Table C.1. Summary Statistics of Share of AI Jobs

		Mean (%)			SD (%)	% >5 AI	Max (%)	Percentiles of All AI Share (%)						
Year	N	All AI	Wave I	Wave II	All AI	Jobs	All AI	P1	P5	P25	P50	P75	P90	P99
2010	4,046	0.20	0.12	0.00	1.37	9.1	50.00	0.00	0.00	0.00	0.00	0.05	0.35	3.30
2011	4,038	0.21	0.13	0.01	1.06	10.6	33.33	0.00	0.00	0.00	0.00	0.07	0.39	3.78
2012	4,085	0.26	0.18	0.01	1.62	12.1	50.00	0.00	0.00	0.00	0.00	0.09	0.46	4.35
2013	4,271	0.31	0.22	0.01	2.07	13.8	100.00	0.00	0.00	0.00	0.00	0.12	0.55	5.00
2014	4,335	0.34	0.22	0.01	1.63	16.2	50.00	0.00	0.00	0.00	0.00	0.16	0.67	4.97
2015	4,254	0.40	0.27	0.01	2.16	18.2	100.00	0.00	0.00	0.00	0.00	0.22	0.85	5.93
2016	4,208	0.48	0.31	0.01	2.71	20.9	100.00	0.00	0.00	0.00	0.00	0.28	0.96	5.88
2017	4,153	0.59	0.39	0.02	2.93	24.2	100.00	0.00	0.00	0.00	0.00	0.39	1.23	7.51
2018	4,156	0.77	0.51	0.03	3.96	27.1	100.00	0.00	0.00	0.00	0.05	0.49	1.50	9.09
2019	4,100	0.81	0.51	0.04	3.29	29.3	100.00	0.00	0.00	0.00	0.10	0.60	1.75	9.70
2020	4,364	0.96	0.56	0.07	3.22	30.2	60.09	0.00	0.00	0.00	0.11	0.69	2.13	13.79
2021	4,589	1.09	0.58	0.09	3.30	31.6	54.79	0.00	0.00	0.00	0.15	0.81	2.60	15.85
2022	4,487	1.18	0.55	0.12	3.58	34.1	100.00	0.00	0.00	0.00	0.20	0.94	2.81	15.58
2023	4,263	1.31	0.55	0.19	3.75	36.1	100.00	0.00	0.00	0.00	0.27	1.08	3.15	16.79
2024	4,022	1.46	0.57	0.25	3.71	38.6	71.69	0.00	0.00	0.00	0.33	1.22	3.64	18.18
All	84,373	0.56	0.31	0.04	2.57	19.3	100.00	0.00	0.00	0.00	0.00	0.27	1.14	9.00

This table reports annual cross-sectional summary statistics for firm-level AI-related U.S. job shares from 2010 to 2024 in the U.S. public-firm sample. AI share is defined as AI-related U.S. jobs divided by total U.S. jobs at the firm and is expressed in percent. Wave I AI includes machine learning (ML), natural language processing (NLP), and computer vision (CV) keywords; Wave II AI includes generative AI (Gen AI) and agentic AI keywords. Wave I and Wave II categories are not mutually exclusive. Mean and SD are computed from the cross-sectional distribution of firm-level shares in each year. The column % $>$ 5 AI jobs reports the share of firms with more than five AI-related U.S. jobs. Percentiles are computed from the cross-sectional distribution of All AI share. The All row pools all firm-years. Firm-years are restricted to those with positive U.S. employment in Revelio and a coverage ratio between 0.05 and 3, where the coverage ratio is defined as total global Revelio workers divided by Compustat employment for each firm-year.

Table C.2. Sample Construction: Effect of Filters on AI Share in 2024

Step	Filter	Firms	Mean (Share of AI Jobs)
0	Compustat–Revelio match + coverage ratio $\in (0.05, 3)$	4,022	1.46%
1	+ U.S. headquartered	3,270	1.39%
2	+ Revelio U.S. workers ≥ 50	2,748	1.42%
3	+ Majority U.S. workers (>50%)	2,262	1.20%
4	+ Compustat employment ≥ 50	2,215	1.18%
5	+ Exclude IT sectors (NAICS 51, 54)	1,923	0.77%
6	+ Drop NAICS 334413 (semiconductor)	1,894	0.73%
7	+ Winsorize Share of AI jobs at 1/99%	1,894	0.64%
8	+ Long-difference panel (2018 & 2024)	1,344	0.56%
9	+ Regression sample (controls non-missing)	1,321	0.55%

This table reports the effect of our sample filters on the number of firms and the mean firm-level share of AI-related U.S. jobs (AI-related jobs / total U.S. jobs) in 2024. Steps 0–5 apply the basic sample restrictions (coverage, U.S. headquartered, U.S. employment, exclude Information / Professional sectors). Step 6 drops semiconductor manufacturers (NAICS 334413), Step 7 winsorizes the AI share at 1/99%, Step 8 keeps firms observed in both 2018 and 2024 (the long-difference panel), and Step 9 restricts to the exact regression sample (firms with all controls non-missing).

Table C.3. Summary Statistics

Panel A: 2010–2018

Variable	N	mean	sd	p1	p5	p10	p25	p50	p75	p90	p95	p99
Change in share of AI jobs, 2010–2018	1222	.0021	.004	-.0031	-.00019	0	0	.00075	.0028	.006	.0091	.024
Change in log sales, 2010–2018	1222	.48	.67	-1.4	-.53	-.24	.12	.44	.8	1.3	1.7	2.7
Change in log employment, 2010–2018	1222	.29	.61	-1.7	-.69	-.37	-.022	.27	.6	1	1.3	2.2
Change in log sales per worker, 2010–2018	1222	.19	.43	-.74	-.37	-.22	-.0083	.16	.33	.61	.89	1.5
Change in revenue TFP, 2010–2018	1091	.039	.34	-.98	-.48	-.32	-.12	.018	.19	.39	.58	1.4
Share of AI jobs, 2010	1224	.00088	.0028	0	0	0	0	0	.00058	.0022	.0044	.013
Have at least 5 AI jobs, 2010	1222	.1	.31	0	0	0	0	0	0	1	1	1
Share of AI jobs, 2018	1222	.003	.0055	0	0	0	0	.0011	.0034	.0078	.012	.033
Have at least 5 AI jobs, 2018	1222	.35	.48	0	0	0	0	0	1	1	1	1

Panel B: 2018–2024

Variable	N	mean	sd	p1	p5	p10	p25	p50	p75	p90	p95	p99
Change in share of AI jobs, 2018–2024	1321	.0028	.0047	-.0099	-.0019	-.00017	0	.0015	.0044	.0087	.013	.02
Change in log sales, 2018–2024	1321	.44	.63	-.98	-.41	-.17	.1	.36	.67	1.1	1.5	3.1
Change in log employment, 2018–2024	1321	.16	.48	-1.2	-.57	-.36	-.09	.11	.37	.73	1	1.8
Change in log sales per worker, 2018–2024	1321	.29	.45	-.96	-.27	-.11	.083	.25	.45	.71	.99	2
Change in revenue TFP, 2018–2024	1173	.024	.39	-1.4	-.51	-.32	-.14	.0036	.17	.38	.59	1.7
Share of AI jobs, 2018	1321	.0029	.0054	0	0	0	0	.00094	.0034	.0076	.012	.031
Have at least 5 AI jobs, 2018	1321	.31	.46	0	0	0	0	0	1	1	1	1
Share of AI jobs, 2024	1321	.0055	.0073	0	0	0	.00057	.0032	.0071	.014	.02	.038
Have at least 5 AI jobs, 2024	1321	.48	.5	0	0	0	0	0	1	1	1	1

This table reports summary statistics for the sample of U.S. public firms (in non-tech sectors) in our baseline regressions (including 1,222 firms in the regressions from 2010 to 2018 and 1,321 firms in the regressions from 2018 to 2024). For each variable, we report the number of observations, the mean, the standard deviation, the median, and 1st, 5th, 10th, 25th, 75th, 90th, 95th, and 99th percentiles.

Table C.4. AI Investments and Firm Growth

Panel A: 2010–2018

	Δ Log Sales		Δ Log Employment	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.092**	0.075**	0.093***	0.074*
	(0.038)	(0.038)	(0.029)	(0.039)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.127	0.264	0.114	0.212
Observations	1,222	1,222	1,222	1,222

Panel B: 2018–2024

	Δ Log Sales		Δ Log Employment	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.068***	0.067**	0.020	0.005
	(0.026)	(0.034)	(0.018)	(0.026)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.162	0.211	0.055	0.109
Observations	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of firm growth on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2010 to 2018 in Panel A and from 2018 to 2024 in Panel B. We consider two measures of firm growth: changes in log sales (columns 1 and 2) and changes in log employment (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of each period (2010 in Panel A and 2018 in Panel B): firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the period (2010 in Panel A and 2018 in Panel B). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.5. AI Investments and Firm Growth Based on the Cognism Sample

Panel A: 2010–2018

	Δ Log Sales		Δ Log Employment	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.145***	0.140***	0.139***	0.129***
	(0.042)	(0.047)	(0.033)	(0.049)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.146	0.298	0.111	0.238
Observations	627	627	627	627

Panel B: 2018–2024

	Δ Log Sales		Δ Log Employment	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.082**	0.077	0.014	-0.007
	(0.038)	(0.049)	(0.020)	(0.029)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.217	0.284	0.102	0.157
Observations	493	493	493	493

This table reports the coefficients from long-differences regressions of firm growth on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors) that are covered in the Cognism dataset. All changes in firm outcomes and share of AI jobs are measured from 2010 to 2018 in Panel A and from 2018 to 2024 in Panel B. We consider two measures of firm growth: changes in log sales (columns 1 and 2) and changes in log employment (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of each period (2010 in Panel A and 2018 in Panel B): firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the period (2010 in Panel A and 2018 in Panel B). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.6. AI Investments and Employment Growth in Revelio

	Δ Log Number of Workers in Revelio		Δ Log Number of Non-AI Workers in Revelio	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.011	0.004	0.006	-0.002
	(0.011)	(0.014)	(0.011)	(0.014)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.135	0.210	0.130	0.205
Observations	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of changes in the number of workers in Revelio on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2018 to 2024. The dependent variables are log change in the number of workers in Revelio in columns 1 and 2 and log change in the number of non-AI workers in Revelio in columns 3 and 4. The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. Columns 2 and 4 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. All specifications control for 2-digit NAICS industry sector fixed effects. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.7. AI Investments and Productivity Growth across Sectors

	Manufacturing		Wholesale & Retail		Finance		Other Non-tech		Tech	
	Sales per Worker	TFP	Sales per Worker	TFP	Sales per Worker	TFP	Sales per Worker	TFP	Sales per Worker	TFP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ Share AI Jobs	0.046**	0.059**	0.135**	0.123**	0.068	0.102	0.098***	0.087***	0.029	0.034*
	(0.019)	(0.024)	(0.065)	(0.055)	(0.069)	(0.072)	(0.033)	(0.031)	(0.019)	(0.019)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.239	0.218	0.243	0.216	0.132	0.217	0.276	0.188	0.246	0.197
Observations	505	474	146	141	273	227	397	331	176	161

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments, separately for each sector: Manufacturing (columns 1 and 2), Wholesale & Retail (columns 3 and 4), Finance (columns 5 and 6), Other Non-tech (columns 7 and 8), and Tech (Information and Professional Services, columns 9 and 10). The sample here includes tech firms, reported separately in the last two columns. For each sector, we report results for changes in log sales per worker and changes in revenue TFP. All changes in productivity and share of AI jobs are measured from 2018 to 2024. The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects and include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.8. AI Investments and Subsequent Productivity Growth

	Δ Log Sales per Worker, t to $t+4$			Δ TFP, t to $t+4$		
	t =2018	t =2019	t =2020	t =2018	t =2019	t =2020
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs, $t-4$ to t	-0.005	0.012	0.017*	-0.006	0.018	0.030***
	(0.013)	(0.013)	(0.011)	(0.017)	(0.013)	(0.011)
Industry FE	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.166	0.089	0.249	0.070	0.093	0.140
Observations	1,215	1,205	1,179	1,127	1,120	1,104

This table reports the coefficients from OLS regressions of changes in firm productivity from year t to $t+4$ on the firm-level changes in AI investments in U.S. public firms (in non-tech sectors) from year $t-4$ to t, where year t is 2018, 2019, or 2020. We consider two measures of productivity: changes in log sales per worker (columns 1 to 3) and changes in revenue TFP (columns 4 to 6). The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions are weighted by the number of jobs in Revelio in year t. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.9. AI Investments and Previous Productivity Growth

	Δ Log Sales per Worker, $t-4$ to t			Δ TFP, $t-4$ to t		
	t=2014	t=2015	t=2016	t=2014	t=2015	t=2016
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs, t to $t+4$	-0.015	0.006	-0.001	0.004	0.015	0.002
	(0.019)	(0.013)	(0.015)	(0.014)	(0.010)	(0.018)
Industry FE	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.169	0.066	0.054	0.082	0.040	0.029
Observations	1,208	1,188	1,187	1,086	1,067	1,063

This table reports the coefficients from OLS regressions of changes in firm productivity from year $t-4$ to t on the firm-level changes in AI investments in U.S. public firms (in non-tech sectors) from year t to $t+4$, where year t is 2014, 2015, or 2016. We consider two measures of productivity: changes in log sales per worker (columns 1 to 3) and changes in revenue TFP (columns 4 to 6). The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions are weighted by the number of jobs in Revelio in year t. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.10. AI Investments and Productivity Growth Using Alternative TFP Measure

	Δ TFP (İmrohoroğlu and Tüzel)		Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	0.063*	0.066**	0.053**	0.077***	0.056***	0.078***
	(0.033)	(0.032)	(0.025)	(0.025)	(0.019)	(0.021)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.078	0.179	0.150	0.238	0.114	0.208
Observations	709	709	709	709	709	709

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider three measures of productivity: changes in the TFP measure from Imrohoroğlu and Tüzel (2014) (columns 1 and 2), changes in log sales per worker (columns 3 and 4) and changes in the baseline revenue TFP measure (columns 5 and 6). In all columns, the sample consists of firms with a non-missing 2018–2024 change in the İmrohoroğlu and Tüzel TFP measure. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 4, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.11. AI Investments and Productivity Growth Excluding Entrants After 2010

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.039**	0.049**	0.053***	0.064***
	(0.017)	(0.019)	(0.013)	(0.015)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.237	0.324	0.105	0.216
Observations	881	881	818	818

This table reports the coefficients from long-differences regressions of changes in firm productivity on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), excluding firms that entered after 2010. All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.12. AI Investments and Productivity Growth Using Share of Wage Bill of AI Workers

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share of Wage Bill of AI Workers	0.043**	0.053**	0.053***	0.063***
	(0.019)	(0.021)	(0.019)	(0.021)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.170	0.216	0.055	0.132
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of wage bill of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is the growth in the share of the wage bill going to AI-related jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.13. AI Investments and Productivity Growth Using Global Share of AI Jobs

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.044**	0.058**	0.057***	0.070***
	(0.021)	(0.026)	(0.021)	(0.026)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.173	0.221	0.063	0.143
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in global AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the global share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.14. AI Investments and Productivity Growth Using 2025 Revelio Labs Vintage

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.051**	0.063***	0.065***	0.079***
	(0.022)	(0.024)	(0.020)	(0.023)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.177	0.222	0.070	0.148
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the U.S. share of AI jobs, based on the Revelio Labs vintage delivered in October 2025 and standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.15. AI Investments and Productivity Growth Using Year-Specific AI Skills

Panel A: 2010–2018

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs (year-specific)	-0.004	-0.000	0.014	0.011
	(0.014)	(0.015)	(0.012)	(0.011)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.089	0.171	0.047	0.150
Observations	1,222	1,222	1,091	1,091

Panel B: 2018–2024

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs (year-specific)	0.049**	0.064**	0.062***	0.078***
	(0.022)	(0.026)	(0.021)	(0.025)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.179	0.229	0.075	0.158
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity on changes in AI investments in U.S. public firms in non-tech sectors. All changes in productivity and the share of AI jobs are measured from 2010 to 2018 in Panel A and from 2018 to 2024 in Panel B. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is the change in the share of AI jobs using the year-specific measure constructed from Revelio Labs resume data, standardized to mean zero and standard deviation of one. To construct the exact-year measure, we build a separate AI keyword list for each year: AI-related skills are identified from Lightcast job postings using the same method as in Appendix A.2.1, and a skill's keywords enter year t's list only if the skill appears in postings in year t. A job is then classified as an AI job in year t only if its title or description contains a keyword from year t's list. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include baseline controls measured at the start of each period (2010 in Panel A and 2018 in Panel B): firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the period (2010 in Panel A and 2018 in Panel B). Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.16. AI Investments and Productivity Growth Excluding Subsidiaries' AI Investments

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.049**	0.062**	0.058***	0.071***
	(0.021)	(0.024)	(0.020)	(0.024)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.179	0.228	0.068	0.151
Observations	1,309	1,309	1,162	1,162

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs, excluding jobs in subsidiary firms, standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.17. AI Investments and Balance-Sheet Intangibles

	Δ Balance-Sheet Intangibles/Total Assets		Δ Acquired Intangibles/Total Assets	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.006	0.002	-0.003	-0.005
	(0.004)	(0.004)	(0.003)	(0.003)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.099	0.149	0.102	0.156
Observations	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of changes in firms' balance-sheet intangibles on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in intangibles and share of AI jobs are measured from 2018 to 2024. In columns 1 and 2, the dependent variable is the change in balance-sheet intangibles (*intan* from Compustat) normalized by total assets; in columns 3 and 4, the dependent variable is the change in acquired intangibles (*acqintan* from Compustat) normalized by total assets. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.18. AI Investments and Productivity Growth Excluding Magnificent 7 Stocks

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.061**	0.073**	0.070***	0.087***
	(0.026)	(0.029)	(0.026)	(0.027)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.193	0.238	0.083	0.160
Observations	1,319	1,319	1,171	1,171

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors) and excluding Magnificent 7 stocks (Apple, Microsoft, Amazon, Alphabet, Meta, Tesla, and Nvidia). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.19. AI Investments and Productivity Growth Controlling for State Fixed Effects and Initial Productivity

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.056***	0.052***	0.066***	0.054***
	(0.020)	(0.018)	(0.018)	(0.018)
TFP in 2010		-0.795***		-0.821***
		(0.133)		(0.117)
Log Sales per Worker in 2010		-0.068**		-0.019
		(0.031)		(0.031)
Industry FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Adj R-Squared	0.295	0.453	0.240	0.415
Observations	1,320	1,179	1,172	1,172

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), controlling for state fixed effects and initial productivity levels. All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. Columns 2 and 4 additionally control for initial TFP in 2010 and log sales per worker in 2010. All specifications control for 2-digit NAICS industry sector fixed effects, state fixed effects, and include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.20. AI Investments and Productivity Growth Controlling for Pre-trends

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.059**	0.067**	0.077***	0.081***
	(0.029)	(0.029)	(0.027)	(0.027)
Industry FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
Pre-trend in sales and employment	Y	Y	Y	Y
Pre-trend in TFP	N	Y	N	Y
Observations	1,007	879	928	876

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), controlling for pre-trends in firm outcomes. All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for pre-trends in sales and employment from 2010 to 2018. Columns 2 and 4 additionally control for pre-trends in TFP. All specifications control for 2-digit NAICS industry sector fixed effects and include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.21. AI Investments and Productivity Growth Controlling for Detailed Industry Fixed Effects

	Δ Log Sales per Worker				Δ TFP			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Share AI Jobs	0.062**	0.070**	0.078*	0.086*	0.075***	0.092***	0.087**	0.095**
	(0.026)	(0.030)	(0.041)	(0.046)	(0.024)	(0.028)	(0.041)	(0.044)
Industry FE	Y	N	N	N	Y	N	N	N
NAICS3 FE	N	Y	N	N	N	Y	N	N
NAICS4 FE	N	N	Y	N	N	N	Y	N
NAICS5 FE	N	N	N	Y	N	N	N	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.226	0.273	0.321	0.336	0.153	0.213	0.262	0.255
Observations	1,321	1,311	1,266	1,171	1,173	1,166	1,119	1,031

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), controlling for increasingly detailed industry fixed effects. All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1–4) and changes in revenue TFP (columns 5–8). Each column uses a different level of industry classification: 2-digit NAICS (columns 1 and 5), 3-digit NAICS (columns 2 and 6), 4-digit NAICS (columns 3 and 7), and 5-digit NAICS (columns 4 and 8). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.22. AI Investments and Productivity Growth Controlling for Covid Exposure and Remote Work Suitability

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.065**	0.062**	0.078***	0.074***
	(0.028)	(0.027)	(0.027)	(0.026)
Covid Exposure	-0.029*		-0.032**	
	(0.015)		(0.015)	
Remote Work Suitability Share		-0.012		0.033
		(0.122)		(0.138)
Industry FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
State FE	Y	Y	Y	Y
Adj R-Squared	0.228	0.225	0.163	0.152
Observations	1,138	1,321	1,021	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), controlling for Covid exposure and remote work suitability. All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector, state fixed effects, and include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 1 and 3 additionally control for firms' Covid exposure between 2020 Q2 and Q4 in Hassan et al. (2023), and columns 2 and 4 additionally control for the share of remote-work-suitable jobs (from Dingel and Neiman (2020)) in 2018. Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.23. AI Investments and Productivity Growth Controlling for Non-AI IT Investments

Panel A: Controlling for 2018 Share of Non-AI IT Jobs

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs, 2018–2024	0.040**	0.040**	0.050***	0.063***
	(0.017)	(0.020)	(0.016)	(0.017)
Share of Non-AI IT Jobs, 2018	0.207	0.689	0.279	0.416
	(0.295)	(0.435)	(0.327)	(0.437)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.179	0.237	0.073	0.157
Observations	1,321	1,321	1,173	1,173

Panel B: Controlling for 2018–2024 Change in Share of Non-AI IT Jobs

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs, 2018–2024	0.047**	0.060**	0.061***	0.074***
	(0.022)	(0.026)	(0.020)	(0.025)
Δ Share Non-AI IT Jobs, 2018–2024	0.030*	0.016	0.003	0.014
	(0.018)	(0.018)	(0.021)	(0.016)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.184	0.227	0.070	0.154
Observations	1,321	1,321	1,173	1,173

This table reports the coefficients from long-differences regressions of changes in firm productivity from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in productivity and share of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 and 2) and changes in revenue TFP (columns 3 and 4). The independent variable is growth in the share of AI jobs, standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector. Panel A controls for the share of non-AI IT jobs (defined as non-AI jobs in SOC 15-1200 computer and IT occupations) in 2018, and Panel B controls for 2018–2024 firm-level changes in share of non-AI IT jobs (standardized to mean zero and standard deviation of one). In each panel, columns 2 and 4 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.24. AI Investments and Firm Tangible and Intangible Investments

	Δ Log SG&A		Δ Log R&D		Δ Log PP&E	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	0.051**	0.052**	0.054**	0.025*	-0.012	-0.007
	(0.022)	(0.023)	(0.025)	(0.014)	(0.025)	(0.024)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.103	0.169	0.195	0.285	0.068	0.193
Observations	1,066	1,066	1,321	1,321	1,238	1,238

This table reports the coefficients from long-differences regressions of changes in firms' investments in intangible and tangible capital from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2018 to 2024. We consider three dependent variables: changes in log SG&A (columns 1 and 2), changes in log R&D (columns 3 and 4), and changes in log PP&E (columns 5 and 6). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 4, and 6 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.25. AI Investments and Employment Shares by Worker Education and Seniority

Panel A: Education

	No College Degree		College Degree		Master's Degree		Doctor's Degree	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Share AI Jobs	-0.686***	-0.454***	0.156	-0.045	0.462***	0.438***	0.069	0.062**
	(0.171)	(0.159)	(0.136)	(0.138)	(0.146)	(0.090)	(0.046)	(0.024)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
Adj R-Squared	0.224	0.301	0.218	0.323	0.128	0.273	0.105	0.224
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321

Panel B: Seniority

	Low Seniority		Medium Seniority		High Seniority	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	-1.134***	-1.168***	0.213	0.270	0.922***	0.899***
	(0.160)	(0.236)	(0.154)	(0.197)	(0.077)	(0.078)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.259	0.316	0.146	0.185	0.280	0.311
Observations	1,321	1,321	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of changes in employment shares by education and seniority levels from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the share of employment (in percentages) of workers in each education level in Panel A and in each seniority level in Panel B. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Even-numbered columns also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.26. AI Investments and Employment Growth by Worker Education and Seniority

Panel A: Education

	No College Degree		College Degree		Master's Degree		Doctor's Degree	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Share AI Jobs	-0.014	-0.030*	0.015	-0.000	0.052***	0.041**	0.057***	0.050***
	(0.017)	(0.018)	(0.014)	(0.018)	(0.016)	(0.016)	(0.021)	(0.018)
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
Adj R-Squared	0.069	0.128	0.173	0.253	0.279	0.355	0.382	0.482
Observations	1,321	1,321	1,321	1,321	1,321	1,321	1,321	1,321

Panel B: Seniority

	Low Seniority		Medium Seniority		High Seniority	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	-0.017	-0.023	0.005	0.003	0.029**	0.021
	(0.011)	(0.017)	(0.014)	(0.014)	(0.013)	(0.016)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.078	0.164	0.091	0.142	0.112	0.149
Observations	1,321	1,321	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of changes in log employment by education and seniority levels from 2018 to 2024 on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in log employment of workers in each education level in Panel A and in each seniority level in Panel B. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Even-numbered columns also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.27. AI Investments and Firm Wage Growth

	Δ Log Average Wage		Δ Log Average Wage of Non-AI Workers	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs	0.013***	0.017***	0.011***	0.016***
	(0.004)	(0.004)	(0.004)	(0.004)
Industry FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.193	0.254	0.182	0.242
Observations	1,321	1,321	1,321	1,321

This table reports the coefficients from long-differences regressions of changes in firm wages on contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors). All changes in firm outcomes and share of AI jobs are measured from 2018 to 2024. We consider two measures of wages: changes in log average wage (columns 1 and 2), and changes in log average wage of non-AI workers (columns 3 and 4). The independent variable is growth in the share of AI jobs (based on the resume data), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2 and 4 also include the baseline controls measured at the start of the period: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.28. AI Investments and Subsequent Organization Capital Growth

	Δ Log(Organization Capital/PP&E), t to $t+4$			Δ Log(Organization Capital/Total Assets), t to $t+4$		
	t=2018	t=2019	t=2020	t=2018	t=2019	t=2020
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs, $t-4$ to t	0.052**	0.050**	0.046*	0.083**	0.068**	0.048**
	(0.026)	(0.020)	(0.024)	(0.039)	(0.033)	(0.023)
Industry FE	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.056	0.103	0.098	0.063	0.104	0.093
Observations	841	826	793	1,025	1,018	990

This table reports the coefficients from OLS regressions of changes in organization capital (based on capitalized SG&A) from year t to $t+4$ on the firm-level changes in AI investments in U.S. public firms (in non-tech sectors) from year $t-4$ to t, where year t is 2018, 2019, or 2020. The dependent variables are changes in the natural log of organization capital normalized by PP&E in columns 1 to 3 and changes in the natural log of organization capital normalized by total assets in columns 4 to 6. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions are weighted by the number of jobs in Revelio in year t. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.29. AI Investments and Previous Organization Capital Growth

	Δ Log(Organization Capital/PP&E), $t-4$ to t			Δ Log(Organization Capital/Total Assets), $t-4$ to t		
	t=2014	t=2015	t=2016	t=2014	t=2015	t=2016
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs, t to $t+4$	-0.062	-0.024	-0.040	-0.013	0.022	-0.006
	(0.040)	(0.033)	(0.039)	(0.012)	(0.016)	(0.043)
Industry FE	Y	Y	Y	Y	Y	Y
Adj R-Squared	0.077	0.087	0.114	0.109	0.070	0.065
Observations	848	824	812	1,019	999	992

This table reports the coefficients from OLS regressions of changes in organization capital (based on capitalized SG&A) from year $t-4$ to t on the firm-level changes in AI investments in U.S. public firms (in non-tech sectors) from year t to $t+4$, where year t is 2014, 2015, or 2016. The dependent variables are changes in the natural log of organization capital normalized by PP&E in columns 1 to 3 and changes in the natural log of organization capital normalized by total assets in columns 4 to 6. The independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs), standardized to mean zero and standard deviation of one. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions are weighted by the number of jobs in Revelio in year t. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.30. AI Investments and Organization Capital Growth: IV Estimates

	Δ Log(Organization Capital/PP&E)			Δ Log(Organization Capital/Total Assets)		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	0.318	0.279*	0.303**	0.253	0.189	0.239**
	(0.244)	(0.146)	(0.138)	(0.259)	(0.129)	(0.109)
Industry FE	Y	Y	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y
State FE	N	Y	Y	N	Y	Y
Control Pre-trend	N	N	Y	N	N	Y
F Statistic	6.3	16.6	19.2	6.3	16.5	22.8
Observations	419	408	383	523	514	439

This table shows 2SLS estimates of the effect of changes in AI investments on changes in organization capital for U.S. public firms (in non-tech sectors). All changes in measures of organization capital and the share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the natural log of organization capital normalized by PP&E in columns 1 to 3 and changes in the natural log of organization capital normalized by total assets in columns 4 to 6. The (instrumented) independent variable is growth in the share of AI jobs (based on the resume data from Revelio Labs) from 2018 to 2024, standardized to mean zero and standard deviation of one. The instrument measures ex-ante firm-level exposure to the supply of AI-trained university graduates from universities with historically strong AI research: it is the share of STEM workers in the firm between 2005 and 2010 who graduated from ex-ante AI-strong research universities. The first-stage F-statistics of the instrument are reported for all specifications. All specifications control for 2-digit NAICS industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 20 universities. All columns also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 2, 3, 5, and 6 add state fixed effects, and columns 3 and 6 additionally control for firm-level changes in log sales, log employment, and TFP from 2010 to 2018. Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.31. Organization-Capital-Building AI Jobs and Productivity Growth with Alternative Thresholds

Panel A: Using 0.5 Threshold for Organization-Capital-Building Jobs

	Δ Log Sales per Worker			Δ TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.068***	0.083***	0.089***	0.080***	0.099***	0.100***
	(0.024)	(0.028)	(0.028)	(0.024)	(0.026)	(0.026)
Δ Share of Non-OC-building AI Jobs			-0.014			-0.004
			(0.019)			(0.019)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.191	0.241	0.241	0.084	0.172	0.172
Observations	1,306	1,306	1,306	1,163	1,163	1,163

Panel B: Using 0.7 Threshold for Organization-Capital-Building Jobs

	Δ Log Sales per Worker			Δ TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.064***	0.079***	0.054***	0.076***	0.095***	0.061***
	(0.024)	(0.026)	(0.019)	(0.025)	(0.026)	(0.019)
Δ Share of Non-OC-building AI Jobs			0.040			0.052*
			(0.029)			(0.028)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.178	0.226	0.233	0.062	0.149	0.164
Observations	1,306	1,306	1,306	1,163	1,163	1,163

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in OC-building AI jobs and non-OC-building AI jobs in U.S. public firms (in non-tech sectors) under alternative thresholds for OC-building jobs. All changes in productivity and the shares of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 to 3) and changes in revenue TFP (columns 4 to 6). The independent variables are growth in the share of OC-building AI jobs (normalized by firms' total U.S.-based jobs) in all columns, and growth in the share of AI jobs not related to OC (normalized by firms' total U.S.-based jobs) in columns 3 and 6, which are both standardized to mean zero and standard deviation of one. In Panel A, a job is OC-building if the OC score (based on job description in the Revelio resume data) is at least 0.5. In Panel B, a job is OC-building if the OC score is at least 0.7. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 3, 5, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.32. Organization-Capital-Building AI Jobs and Organization Capital Growth

	Δ Log(Organization Capital/PP&E)			Δ Log(Organization Capital/Total Assets)		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.145***	0.131***	0.115***	0.159***	0.134***	0.103**
	(0.019)	(0.024)	(0.028)	(0.048)	(0.041)	(0.043)
Δ Share of Non-OC-building AI Jobs			0.038			0.065*
			(0.033)			(0.036)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.120	0.146	0.149	0.105	0.129	0.137
Observations	877	877	877	1,092	1,092	1,092

This table reports the coefficients from long-differences regressions of changes in OC on the contemporaneous firm-level changes in OC-building AI jobs and non-OC-building AI jobs in U.S. public firms (in non-tech sectors). All changes in OC and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the natural log of OC normalized by PP&E in columns 1 to 3 and changes in the natural log of OC normalized by total assets in columns 4 to 6. The independent variables are growth in the share of OC-building AI jobs (normalized by firms' total U.S.-based jobs) in all columns, and growth in the share of AI jobs not related to OC (normalized by firms' total U.S.-based jobs) in columns 3 and 6, which are both standardized to mean zero and standard deviation of one. A job is OC-building if the OC score (based on job description in the Revelio resume data) is at least 0.6. All specifications control for 2-digit NAICS industry sector fixed effects. Columns 2, 3, 5, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.33. Organization-Capital-Building AI Jobs and Productivity Growth Controlling for Initial Organization Capital

	Δ Log Sales per Worker			Δ TFP		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.087***	0.110***	0.107***	0.090***	0.115***	0.102***
	(0.025)	(0.028)	(0.027)	(0.023)	(0.025)	(0.025)
Δ Share of Non-OC-building AI Jobs			0.006			0.027
			(0.027)			(0.026)
Log(Organization Capital/PP&E), 2018	-0.010	-0.005	-0.005	-0.027**	-0.020**	-0.020**
	(0.012)	(0.010)	(0.010)	(0.011)	(0.009)	(0.009)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.164	0.232	0.231	0.142	0.217	0.220
Observations	881	881	881	842	842	842

This table reports the coefficients from long-differences regressions of changes in firm productivity on the contemporaneous firm-level changes in OC-building AI jobs and non-OC-building AI jobs in U.S. public firms (in non-tech sectors). All changes in productivity and the shares of AI jobs are measured from 2018 to 2024. We consider two measures of productivity: changes in log sales per worker (columns 1 to 3) and changes in revenue TFP (columns 4 to 6). The independent variables are growth in the share of OC-building AI jobs (normalized by firms' total U.S.-based jobs) in all columns, and growth in the share of AI jobs not related to OC (normalized by firms' total U.S.-based jobs) in columns 3 and 6, which are both standardized to mean zero and standard deviation of one. A job is OC-building if the OC score (based on job description in the Revelio resume data) is at least 0.6. All specifications control for 2-digit NAICS industry sector fixed effects and the natural log of OC normalized by PP&E in 2018. Columns 2, 3, 5, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the period. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.34. Organization-Capital-Building AI Jobs and Organization Capital Growth Controlling for Initial Organization Capital

	Δ Log(Organization Capital/PP&E)			Δ Log(Organization Capital/Total Assets)		
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share of OC-building AI Jobs	0.129***	0.114***	0.109***	0.151***	0.123**	0.110**
	(0.018)	(0.020)	(0.023)	(0.051)	(0.048)	(0.049)
Δ Share of Non-OC-building AI Jobs			0.013			0.032
			(0.032)			(0.037)
Log(Organization Capital/PP&E), 2018	0.077*	0.091**	0.090**	0.092**	0.108***	0.106***
	(0.045)	(0.044)	(0.045)	(0.038)	(0.035)	(0.037)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
Adj R-Squared	0.179	0.215	0.214	0.178	0.211	0.211
Observations	877	877	877	881	881	881

This table reports the coefficients from long-differences regressions of changes in OC on the contemporaneous firm-level changes in OC-building AI jobs and non-OC-building AI jobs in U.S. public firms (in non-tech sectors). All changes in OC and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the natural log of OC normalized by PP&E in columns 1 to 3 and changes in the natural log of OC normalized by total assets in columns 4 to 6. The independent variables are growth in the share of OC-building AI jobs (normalized by firms' total U.S.-based jobs) in all columns, and growth in the share of AI jobs not related to OC (normalized by firms' total U.S.-based jobs) in columns 3 and 6, which are both standardized to mean zero and standard deviation of one. A job is OC-building if the OC score (based on job description in the Revelio resume data) is at least 0.6. All specifications control for 2-digit NAICS industry sector fixed effects and the natural log of OC normalized by PP&E in 2018. Columns 2, 3, 5, and 6 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio at the start of the period. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.35. Heterogeneous Relationship between AI Investments and Organization Capital Growth by Initial Organization Capital

	Δ Log(Organization Capital/PP&E)		Δ Log(Organization Capital/Total Assets)	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs*Below-Median OC	0.153***	0.141***	0.214***	0.193***
	(0.035)	(0.043)	(0.049)	(0.052)
Δ Share AI Jobs*Above-Median OC	0.064***	0.065**	0.024	0.015
	(0.024)	(0.030)	(0.024)	(0.029)
NAICS2*Size tercile*Above-Median OC FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.162	0.188	0.187	0.214
Observations	871	871	875	875

This table reports the coefficients from long-differences regressions of changes in OC (capitalized SG&A) on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), separately for firms with above-median and below-median initial levels of OC. Firms within each 2-digit NAICS industry sector are split at the median level of OC normalized by PP&E in 2018. All changes in OC and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the natural log of OC normalized by PP&E in columns 1 and 2 and changes in the natural log of OC normalized by total assets in columns 3 and 4. The main independent variable is growth in the share of AI jobs, standardized to mean zero and standard deviation of one, interacted with an indicator for above- versus below-median initial OC. All specifications control for fixed effects for each 2-digit NAICS industry × firm size tercile × above-median organization capital cell. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.36. Heterogeneous Relationship between AI Investments and Organization Capital Growth by Average Organization Capital (2014–2018)

	Δ Log(Organization Capital/PP&E)		Δ Log(Organization Capital/Total Assets)	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs*Below-Median OC	0.157***	0.147***	0.218***	0.199***
	(0.034)	(0.043)	(0.048)	(0.053)
Δ Share AI Jobs*Above-Median OC	0.059**	0.060**	0.019	0.017
	(0.024)	(0.029)	(0.025)	(0.028)
NAICS2*Size tercile*Above-Median OC FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.155	0.172	0.180	0.199
Observations	871	871	882	882

This table reports the coefficients from long-differences regressions of changes in OC (capitalized SG&A) on the contemporaneous firm-level changes in AI investments in U.S. public firms (in non-tech sectors), separately for firms with above-median and below-median average OC over 2014–2018. Firms within each 2-digit NAICS industry sector are split at the median level of the average OC normalized by PP&E between 2014 and 2018. All changes in OC and share of AI jobs are measured from 2018 to 2024. The dependent variables are changes in the natural log of OC normalized by PP&E in columns 1 and 2 and changes in the natural log of OC normalized by total assets in columns 3 and 4. The independent variable is growth in the share of AI jobs, standardized to mean zero and standard deviation of one, interacted with an indicator for above- versus below-median average OC. All specifications control for fixed effects for each 2-digit NAICS industry × firm size tercile × above-median organization capital cell. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.37. AI Investments and Pre-Trends in Organization Capital by Initial Organization Capital

	Δ Log(Organization Capital/PP&E), 2012–2018		Δ Log(Organization Capital/Total Assets), 2012–2018	
	(1)	(2)	(3)	(4)
Δ Share AI Jobs*Below-Median OC	0.005	0.043	-0.039	-0.003
	(0.028)	(0.049)	(0.036)	(0.039)
Δ Share AI Jobs*Above-Median OC	-0.004	-0.008	-0.046	-0.018
	(0.024)	(0.038)	(0.065)	(0.070)
NAICS2*Size tercile*Above-Median OC FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Adj R-Squared	0.244	0.316	0.670	0.719
Observations	625	625	625	625

This table reports the coefficients from regressing changes in OC (capitalized SG&A) from 2012 to 2018 on changes in AI investments from 2018 to 2024 in U.S. public firms (in non-tech sectors), separately for firms with above-median and below-median initial levels of OC. Firms within each 2-digit NAICS industry sector are split at the median level of OC normalized by PP&E in 2018. The dependent variables are 2012–2018 changes in the natural log of OC normalized by PP&E in columns 1 and 2 and changes in the natural log of OC normalized by total assets in columns 3 and 4. The independent variable is 2018–2024 growth in the share of AI jobs, standardized to mean zero and standard deviation of one, interacted with an indicator for above- versus below-median initial OC. All specifications control for fixed effects for each 2-digit NAICS industry × firm size tercile × above-median organization capital cell. Columns 2 and 4 also include the baseline controls measured in 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Regressions are weighted by the number of jobs in Revelio in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.38. AI Investments and Industry Productivity Growth and Price Increases

	Δ Log Sales per Worker		Δ TFP		Δ Log PPI	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	0.073***	0.078***	0.040***	0.033***	0.005	0.009
	(0.023)	(0.030)	(0.010)	(0.010)	(0.016)	(0.014)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.217	0.214	0.247	0.268	0.117	0.201
Observations	287	287	287	287	210	210

This table reports the coefficients from industry-level long-differences regressions of changes in industry productivity and prices on contemporaneous changes in industry-level AI investments from 2018 to 2024. All changes in industry outcomes and share of AI jobs are measured from 2018 to 2024. All industry-level variables are calculated for all firms in Compustat (regardless of whether they are in our main regression sample or not). Each observation is a 5-digit NAICS industry, and (as in our main analysis) we exclude tech sectors. The independent variable is the change in the share of AI jobs among U.S. jobs (based on the resume data), standardized to mean zero and standard deviation of one. The dependent variables are the changes in log sales per worker (log of total industry sales divided by total industry employment) in columns 1 and 2, sales-weighted average TFP in columns 3 and 4, and industry-level producer price index (PPI) in columns 5 and 6. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions in columns 2, 4, and 6 also include industry-level controls for log total employment, log total sales, and log average wage in 2018. Regressions are weighted by the total (industry-level) number of Revelio jobs in 2018. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table C.39. AI Investments and Industry Price Increases and Employment Growth Including Non-Public Firms

	Δ Log PPI		Δ Log Emp (Revelio)		Δ Log Emp (QWI)	
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Share AI Jobs	-0.018	-0.009	0.008	-0.011	-0.008	-0.019
	(0.014)	(0.013)	(0.015)	(0.018)	(0.017)	(0.019)
Industry FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Adj R-Squared	0.110	0.389	0.322	0.374	0.183	0.246
Observations	210	181	287	248	284	246

This table reports the coefficients from industry-level long-differences regressions of changes in industry prices and employment on contemporaneous changes in industry-level AI investments from 2018 to 2024. Each observation is a 5-digit NAICS industry, and (as in our main analysis) we exclude tech sectors. All changes in industry outcomes and share of AI jobs are measured from 2018 to 2024. All industry-level variables are calculated for all public and private firms in the Revelio data (regardless of whether they are in our main regression sample or not). The independent variable is the change in the share of AI jobs among U.S. jobs (based on the resume data), standardized to mean zero and standard deviation of one. The dependent variables are industry-level producer price index (PPI) in columns 1 and 2, the changes in log U.S. employment based on Revelio in columns 3 and 4, and log employment based on QWI in columns 5 and 6. All specifications control for 2-digit NAICS industry sector fixed effects. Regressions in columns 2, 4, and 6 also include industry-level controls for log total employment, log total sales, and log average wage in 2018. Regressions are weighted by the total (industry-level) number of Revelio jobs in 2018. Standard errors are robust to heteroskedasticity and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

D. Instrument Construction and Validation

Our instrumental variable strategy builds on Babina et al. (2024) and exploits differences in firms' pre-existing ties to universities with historically strong AI research traditions to generate variation in firm-level AI investment. Building the instrument requires two distinct data inputs. The first is an indicator of each university's pre-period strength in AI research. The second is a measure of the hiring linkages between firms and specific universities prior to the treatment period. For the first component, we categorize universities as either ex-ante strong in AI or not, on the basis of the count of faculty members publishing AI-related work at each institution before 2010. For the second component—firm-university hiring networks—we draw on the Revelio resume data to identify which universities the existing employees of each firm in 2010 had attended.

Data Construction.

To identify universities with pre-existing strength in AI research, we rely on the definition of AI-strong universities from Babina et al. (2024), which provides a detailed description of the methodology. In a nutshell, we use data from Open Academic Graph (OAGv2) to quantify AI-related publications linked to each institution. We then link 659 research institutions—representing 99% of entries—in the Higher Education Research and Development Survey (HERDS) to faculty records in OAGv2. The HERDS data, collected by the National Science Foundation, cover all U.S. universities with at least $150,000 in annual R&D expenditures. For each university matched to the HERDS roster, we collect all OAGv2 publications in each year for which at least one co-author lists an affiliation with that institution. While we are interested specifically in AI-related researchers to build our instrument, we also identify non-AI CS-related researchers to use as a control variable since AI- and CS-research strengths might be highly correlated. We classify researchers as AI and CS researchers based on the share of their publications in AI- or CS-related venues, and calculate the annual proportion of researchers classified as AI researchers and the proportion classified as CS researchers at the university level. A university is classified as AI-strong if, in at least one year between 2005 and 2009, it satisfies either of the following conditions: (i) its count of AI researchers ranks in the top 5% across all universities in that year; (ii) its count of AI researchers ranks in the top 10% and the *share* of AI researchers relative to all other researchers in the OAGv2 data ranks in the top 5% across all universities in that year.

To build the second component of our instrument—firm-level exposure to historically AI-strong universities through pre-existing hiring networks—we turn to our Revelio resume data. These records list the degree-granting institutions associated with every credential on workers' resumes. First, we disambiguate university names in the Revelio data and match them to the HERDS roster. We then define individual i as a graduate of university u if the most recent degree (graduate or undergraduate) reported on individual i's resume is from university u. We classify individual i as a *fresh graduate* of university u in year t if i entered employment at a firm in year t and the job is the first job after graduating from university u. The data offer extensive university coverage: in 2010, for example, 643 out of 659 universities in the HERDS dataset have at least one fresh graduate represented in our resume records. Because firms' recruitment patterns for STEM versus non-STEM workers may differ—a firm might draw economics talent from one

university's policy program and management talent from another's business school—we also construct firm-university hiring networks restricted to STEM employees, which may be more relevant for AI recruitment. We define STEM workers as those holding at least one degree in engineering (e.g., electrical, chemical, mechanical), physical sciences (e.g., mathematics, physics, chemistry, computer science, statistics), or biological sciences (e.g., biology, pharmacology). In our analysis, we exclude firms for which the fraction of STEM workers with missing university information exceeds 50 percent.

We benchmark the coverage of our university graduate data against official statistics and find that our resume records capture a substantial share of U.S. university graduates. Aggregating to the university-year level, we compute the total number of fresh graduates from each institution in each year and compare these figures to the total degrees (bachelor's, master's, and PhDs) conferred annually, as reported in the Integrated Postsecondary Education Data System (IPEDS), which covers all post-secondary institutions in the United States. As of 2012 (the most recent year in our IPEDS extract), our resume data account for an average of 46% of all fresh graduates at each university. The cross-sectional correlation between our fresh graduate counts and the IPEDS figures is also high, at 0.67. Both figures are high given that we only track students who go on to work at U.S. firms.

Lastly, we use the Revelio resume data to compute the annual share of fresh university graduates from each institution who enter AI-related positions, covering the period from 2006 to 2024. Appendix Figure D.1 shows that the share of AI graduates increased from less than 0.5% in the early 2010s to 4% in 2024, suggesting that universities scale up the production of AI-related graduates once commercial demand for AI applications expanded after the mid-2010s.

Instrument Validation.

We begin by testing several core assumptions that underpin the logic of our instrument. Consistent with our central argument, we demonstrate that the growth in AI-trained graduates during the 2010s was far more pronounced at AI-strong universities than at other institutions. Appendix Figure D.2 displays the share of fresh graduates entering AI-related jobs, plotted separately for AI-strong and non-AI-strong universities over the 2006–2024 period. In 2006, the proportion of AI graduates was negligible across the board, remaining well below 1% for both groups. In 2018, the share grew to 4% at AI-strong universities, but remained at around 1% at other universities. Between 2018 and 2024, AI-strong universities saw this share further double to about 8%, while non-AI-strong universities grew the share to only 3%.

We next examine which universities produce more AI graduates. Appendix Table D.1 and Appendix Table D.2 list the top 10 universities with the highest share and number of AI graduates in 2018 and 2024. When measured by the share of graduates trained in AI, elite technical institutions dominate: Caltech, Carnegie Mellon, Stanford, and MIT consistently rank at the top in both 2018 and 2024, with shares rising markedly over time (e.g., Carnegie Mellon from 12% to 17%). When measured by absolute numbers, large research universities lead: Georgia Tech tops both years (421 in 2018, 545 in 2024), followed by Berkeley, Carnegie Mellon, and Columbia, with most schools seeing substantial growth in AI graduate counts.

We next assess whether firm-university hiring networks generate the variation needed

for our identification strategy. Our instrument relies on cross-firm differences in exposure to AI-strong universities, which requires that firms draw their talent from distinct sets of institutions rather than hiring uniformly from the same pool. Empirically, hiring is highly concentrated: the average firm sources 19% of its fresh graduates from a single dominant university, 47% from its top five universities, and 61% from its top ten. By comparison, the single largest university accounts for only 1.6% of the nationwide pool of fresh graduates, the five largest for 7.1%, and the ten largest for 12.9%. Geographic proximity also plays a meaningful role: on average, firms recruit 36% of all fresh graduates, 34% of STEM fresh graduates, and 32% of AI fresh graduates from universities located in the same state as their headquarters. In addition, for the instrument to function, pre-existing firm-university networks must be persistent enough to predict post-period hiring of AI-related workers. In column 1 of Table D.3, we regress the share of fresh graduates recruited from each university after 2018 on the corresponding pre-2010 share, finding a strong positive association that confirms the temporal stability of these networks. Column 2 uses the share of all workers employed at the firm between 2005 and 2010 who graduated from each university as the predictor, with similarly strong results. This persistence extends to AI hiring specifically: columns 3 and 4 show that the universities supplying a firm's overall workforce before 2010 are also strong predictors of the universities from which that firm later recruits AI-related employees. Finally, columns 5 and 6 demonstrate that pre-2010 firm-university hiring networks constructed solely from STEM workers also have considerable predictive power for post-2018 AI recruitment.

One concern is that hiring ties to AI-strong universities also proxy for exposure to technically skilled or elite graduates more generally, in which case the instrument would capture access to skilled labor broadly. In all specifications, we control separately for firms' historical exposure to computer-science-strong and top-20 universities, $\sum_u s_{iu}^{Pre2010} CSstrong_u$ and $\sum_u s_{iu}^{Pre2010} Top20_u$, where $CSstrong_u$ is the average pre-2010 share of (non-AI) CS researchers at university u and $Top20_u$ indicates a top-20 university in the U.S. News & World Report ranking.

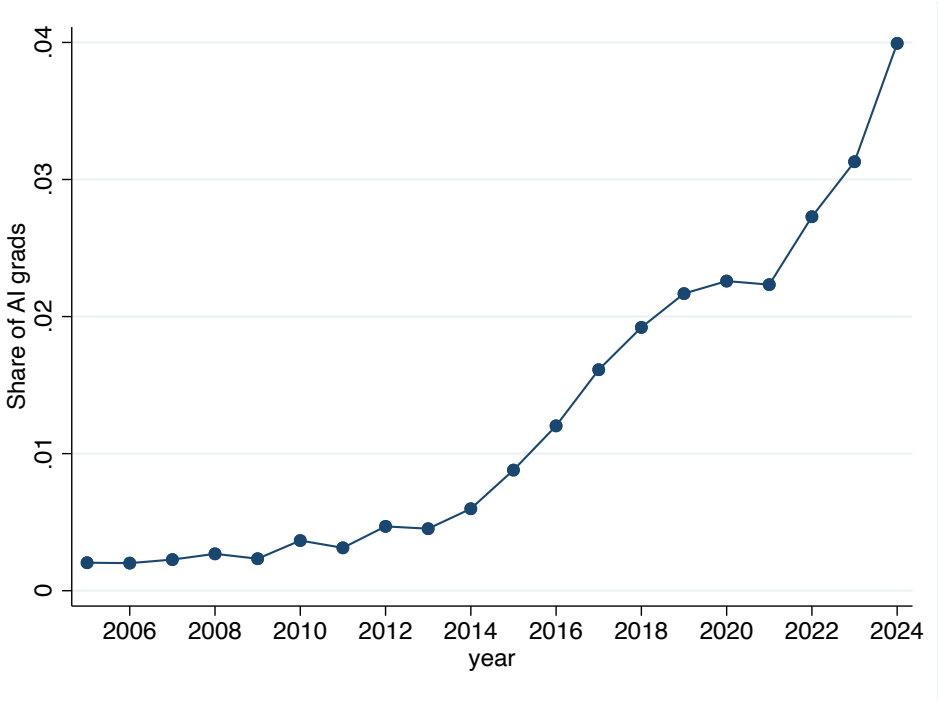
Robustness.

An important identification concern is that if firms anticipated the coming wave of AI demand, they might have proactively strengthened their recruiting ties to AI-strong universities before 2010, rendering the 2010 hiring networks endogenous to future AI recruitment capacity. This scenario appears unlikely, given the absence of both significant corporate interest in AI and meaningful production of AI-related graduates by universities prior to 2010. We also provide direct empirical support: firms connected to AI-strong universities did not disproportionately increase their hiring from those institutions between 2005 and 2010. Appendix Table D.5 confirms that there is no significant association between changes in the share of fresh graduates from AI-strong universities during the pre-period (2005–2010) and our instrument.

Appendix Table D.6 demonstrates that firms with greater exposure to AI-strong universities did not experience faster growth in productivity prior to 2010, lending support to the exclusion restriction that pre-existing ties to AI-strong universities affect firm performance only through AI investments made after 2010. Panel B further shows that exposure to AI-strong universities is not associated with higher productivity growth over 2010–2018,

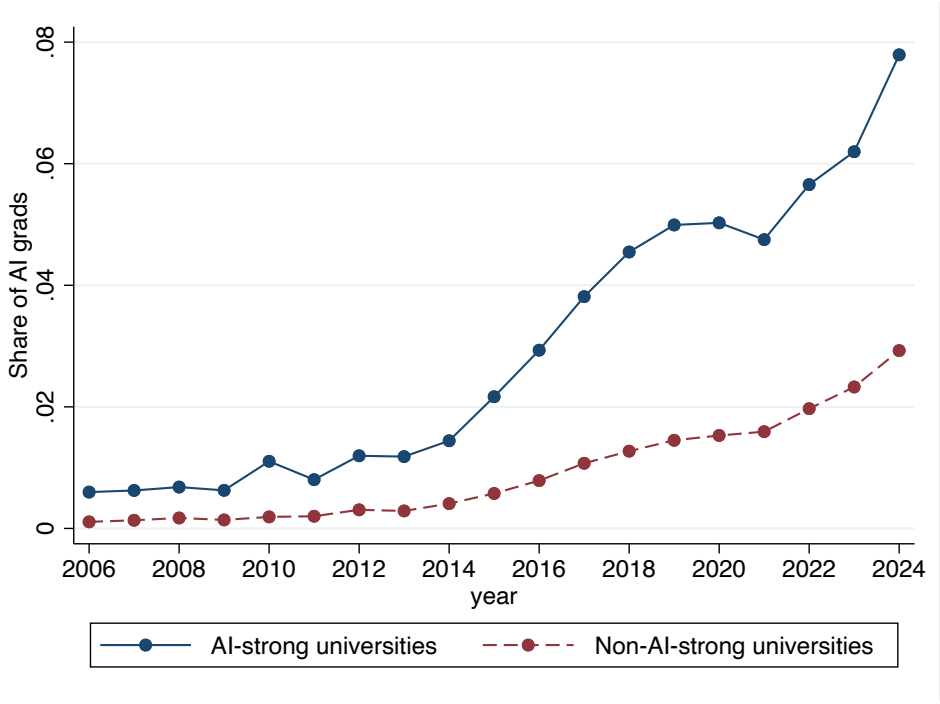
consistent with the lagged effects of AI investment on productivity.

Figure D.1. Share of AI-trained Fresh Graduates Over Time



This figure plots the share of AI-trained fresh graduates out of all fresh graduates from 2005 to 2024 in the U.S. public-firm sample. We define an individual i as a fresh graduate from university u in year t if individual i joined a firm in year t and got the most recent degree from university u and the job is the first job after graduation. An individual is considered an AI-trained fresh graduate in year t if the individual is a fresh graduate in year t and that individual's first job after graduation is an AI-related job. AI-related jobs are defined based on the methodology described in Section 2.1 and used throughout the paper.

Figure D.2. Time Series of the Share of AI-trained Fresh Graduates from Ex-ante AI-strong Universities and Other Universities



This figure plots the share of AI-trained fresh graduates out of all fresh graduates from 2006 to 2024 in the U.S. public-firm sample, separately for ex-ante AI-strong universities and non-AI-strong universities. We define a university as an AI-strong university if it satisfies one of the following two criteria in at least one year between 2005 and 2009: (i) the number of AI researchers is in the top 5% of the distribution across all universities in a given year; (ii) the number of AI researchers is in the top 10% of the distribution, and the share of AI researchers is in the top 5% of the distribution across all universities in a given year. We define an individual i as a fresh graduate from university u in year t if individual i joined a firm in year t and got the most recent degree from university u and the job is the first job after graduation. An individual is considered an AI-trained fresh graduate in year t if the individual is a fresh graduate in year t and that individual's first job after graduation is an AI-related job. AI-related jobs are defined based on the methodology described in Section 2.1 and used throughout the paper.

Table D.1. Top 10 Universities by Share of AI Graduates

	2018		2024	
Rank	**University**	**Share**	**University**	**Share**
1	California Institute of Technology	0.119	CUNY Graduate Center	0.176
2	Carnegie Mellon University	0.117	Carnegie Mellon University	0.170
3	Stanford University	0.097	Rockefeller University	0.167
4	Massachusetts Institute of Technology	0.097	Stanford University	0.143
5	Harvey Mudd College	0.094	Claremont Graduate University	0.133
6	Georgia Institute of Technology	0.068	Massachusetts Institute of Technology	0.132
7	CUNY Graduate Center	0.067	California Institute of Technology	0.121
8	Princeton University	0.056	Illinois Institute of Technology	0.108
9	Rice University	0.055	Georgia Institute of Technology	0.102
10	Williams College	0.050	Columbia University	0.097

This table lists the top 10 universities with the highest share of AI-trained fresh graduates in 2018 and 2024 in the U.S. public-firm sample. An individual is considered an AI-trained fresh graduate in year t if the individual is a fresh graduate in year t and that individual's first job after graduation is an AI-related job. AI-related jobs are defined based on the methodology described in Section 2.1 and used throughout the paper.

Table D.2. Top 10 Universities by Number of AI Graduates

	2018		2024	
Rank	**University**	**Count**	**University**	**Count**
1	Georgia Institute of Technology	421	Georgia Institute of Technology	545
2	Carnegie Mellon University	353	University of California, Berkeley	468
3	University of California, Berkeley	306	Carnegie Mellon University	430
4	Stanford University	292	Northeastern University	412
5	University of Southern California	264	Columbia University	389
6	Columbia University	250	University of Southern California	370
7	New York University	224	New York University	352
8	University of Illinois at Urbana-Champaign	214	University of Illinois at Urbana-Champaign	349
9	University of California, San Diego	177	University of California, San Diego	344
10	Massachusetts Institute of Technology	167	Stanford University	290

This table lists the top 10 universities with the largest number of AI-trained fresh graduates in 2018 and 2024 in the U.S. public-firm sample. An individual is considered an AI-trained fresh graduate in year t if the individual is a fresh graduate in year t and that individual's first job after graduation is an AI-related job. AI-related jobs are defined based on the methodology described in Section 2.1 and used throughout the paper.

Table D.3. Persistence of Firm-University Hiring Networks

	Share of Post-2018 Hires		Share of Post-2018 AI Hires			
	(1)	(2)	(3)	(4)	(5)	(6)
Share of Pre-2010 Hires	0.583***		0.542***			
	(0.021)		(0.040)			
Share of Pre-2010 Workers		0.636***		0.586***		
		(0.023)		(0.050)		
Share of Pre-2010 STEM Hires					0.172***	
					(0.020)	
Share of Pre-2010 STEM Workers						0.292***
						(0.032)
Firm FE	Y	Y	Y	Y	Y	Y
University FE	Y	Y	Y	Y	Y	Y
Observations	118,897	175,419	108,763	157,584	108,651	157,524

This table reports the coefficients from regressing the share of each firm's fresh graduates hired from each university after 2018 on the pre-2010 firm-university network. Each observation is a firm-university pair. The dependent variable, constructed using the Revelio resume data, is the share of all fresh graduates hired from each university after 2018 in columns 1 and 2 and the share of AI-trained fresh graduates hired from each university after 2018 in columns 3–6. In columns 1 and 3, the independent variable is the share of all fresh graduates hired from each university between 2005 and 2010. We define an individual i as a fresh graduate from university u in year t if individual i joined a firm in year t and got the most recent degree from university u and the job is the first job after graduation. In columns 2 and 4, the independent variable is the share of all workers in the firm between 2005 and 2010 who graduated from each university. In column 5, the independent variable is the share of all STEM fresh graduates hired from each university between 2005 and 2010. We define STEM workers as employees who have at least one degree with a major in either engineering (e.g., electrical, chemical, mechanical), physical sciences (e.g., math, physics, chemistry, computer science, statistics), or biological sciences (e.g., biology, pharmacology). In column 6, the independent variable is the share of STEM workers in the firm between 2005 and 2010 who graduated from each university. All columns control for firm fixed effects and university fixed effects. Standard errors are clustered at the university level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table D.4. First Stage of the Instrument

	Δ Share of AI Jobs		
	(1)	(2)	(3)
Instrument	0.362***	0.498***	0.687***
	(0.114)	(0.118)	(0.134)
Industry FE	Y	Y	Y
Baseline Controls	Y	Y	Y
State FE	N	Y	Y
Control Pre-trend	N	N	Y
F Statistic	10.1	17.9	26.2
Observations	633	627	531

This table reports the first stage of the instrument, where we regress firm-level changes in the share of AI jobs over 2018–2024 on the instrument, which measures ex-ante firm-level exposure to the supply of AI-trained university graduates from universities with historically strong AI research. The independent variable is the share of STEM workers in the firm between 2005 and 2010 who graduated from ex-ante AI-strong research universities. The dependent variable is the resume-based measure of the growth in the share of AI jobs from 2018 to 2024. Both the independent variable and the dependent variable are standardized to mean zero and standard deviation of one. All specifications control for industry sector fixed effects, ex-ante firm-level exposure to universities that are historically strong in CS research as well as top 20 universities, and the baseline controls measured as of 2018: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 2 and 3 add state fixed effects. Column 3 additionally controls for firm-level pre-trends: changes in log sales, log employment, and TFP from 2010 to 2018. Regressions are weighted by the number of Revelio jobs in 2018. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. The first-stage F-statistics of the instrument are reported for all specifications. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table D.5. Changes in Hiring from Ex-ante AI-strong Universities during the Pre-period

	Δ Share of Fresh Graduates Hired from AI-Strong Universities 2005–2010		
	(1)	(2)	(3)
Instrument	0.020	0.035	0.016
	(0.029)	(0.038)	(0.036)
Industry FE	Y	Y	Y
Baseline Controls	Y	Y	Y
State FE	N	Y	Y
Control Pre-trend	N	N	Y
Adj R-Squared	0.047	0.070	0.164
Observations	470	463	378

This table reports the coefficients from regressing the change in the share of fresh graduates from AI-strong universities from 2005 to 2010 on the instrument (the share of STEM workers in the firm between 2005 and 2010 who graduated from AI-strong universities). The independent variable is standardized to mean zero and standard deviation of one. All columns control for ex-ante exposure to universities that are strong in CS research and top 20 universities, industry sector fixed effects, and the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 2 and 3 add state fixed effects. Column 3 additionally controls for firm-level pre-trends: changes in log sales, log employment, and TFP from 2002 to 2010. Regressions are weighted by the number of Revelio jobs in 2010. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table D.6. Firm Connections to AI-strong Universities and Productivity Growth in Earlier Periods

Panel A: 2002–2010

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Instrument	0.021	-0.016	0.003	-0.021
	(0.032)	(0.030)	(0.034)	(0.036)
Industry FE	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y
State FE	N	Y	N	Y
Adj R-Squared	0.339	0.409	0.289	0.341
Observations	643	639	550	545

Panel B: 2010–2018

	Δ Log Sales per Worker		Δ TFP	
	(1)	(2)	(3)	(4)
Instrument	-0.011	0.012	-0.034	-0.019
	(0.032)	(0.028)	(0.032)	(0.031)
Industry FE	Y	Y	Y	Y
Baseline Controls	Y	Y	Y	Y
State FE	N	Y	N	Y
Adj R-Squared	0.278	0.292	0.243	0.271
Observations	808	806	714	711

This table reports the coefficients from regressions of firm productivity growth in earlier periods on the instrument—ex-ante firm-level exposure to AI-related graduates from AI-strong universities. The instrument is standardized to mean zero and standard deviation of one. The dependent variables are changes in firm productivity from 2002 to 2010 in Panel A and from 2010 to 2018 in Panel B. In each panel, we consider changes in log sales per worker in columns 1 and 2, and changes in TFP in columns 3 and 4. All specifications control for industry sector fixed effects and ex-ante exposure to universities that are strong in computer science research as well as top 20 universities. All columns also include the baseline controls all measured as of 2010: firm-level characteristics (log sales, cash/assets, R&D/sales, markup, and log number of workers in Revelio), log industry wage, and characteristics of the commuting zones where the firms are located (log average wage, log population, exposure to Chinese import competition, the share of foreign-born workers, the share of routine workers, the share of workers in manufacturing, finance, and IT-related occupations, the share of college-educated workers, and the share of female workers). Columns 2 and 4 additionally control for state fixed effects. Regressions are weighted by the number of Revelio jobs in 2010. Standard errors are clustered at the 5-digit NAICS industry level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

E. A Model of AI and Organization Capital

This appendix sets out a way of thinking about how AI affects organization capital and productivity growth. We extend the firm life-cycle model of Atkeson and Kehoe (2005) to separate two state variables inside a firm: *idiosyncratic productivity* A, a reduced-form advantage following their stochastic life-cycle process with declining drift, and *organization capital* B, the durable documented asset a firm builds out of running its operations. Measured productivity is $\tilde{A} = AB^{\psi}$. AI is modeled as a codification technology: it raises the rate χ at which operations are converted into organization capital, leaving the idiosyncratic process untouched. Three conclusions follow.

1. **A permanent but gradual level effect.** Adoption raises measured productivity permanently by $\psi \log(\chi'/\chi)$ and permanently shifts the composition of firm assets toward the durable stock. The effect is a level shift, not a change in growth: firms do not grow faster forever, they converge to a higher path.

2. **Slow adjustment, with no frictions required.** The post-adoption gap in organization capital and productivity follows an exact first-order recursion whose convergence rate is the depreciation rate δ_B plus the organization-capital growth of untreated firms. Plausible depreciation rates imply half-lives of five years or more. Responses that take many years to build are therefore not evidence of adjustment costs or implementation lags—they are what a stock-flow system delivers when stocks are large relative to flows.

3. **Heterogeneity runs through the organization capital ratio.** Both the size of the impact response and the speed of convergence depend on the firm's state only through its pre-adoption organization capital ratio $r = B/A$. Firms with a low level of codified knowledge stock respond most and fastest, which we test empirically.

Introduction

Atkeson and Kehoe (2005) propose that the value of a firm resides in its *organization capital*: the built-up, firm-specific expertise that makes an established producer more productive than a new one. In their model this expertise is a pair (A, s)—a specific productivity and an age—evolving stochastically with an age-dependent drift and lost permanently when the firm closes.

Their framework treats the firm's accumulated advantage as a single object. We split it in two. *Idiosyncratic productivity* is the reduced-form state governing how large a firm is relative to its competitors. It is deliberately agnostic about mechanism: it may reflect the tacit knowledge its workforce holds but has not written down, its technical efficiency, product appeal, or an accumulated customer base—Foster et al. (2016) show that the slow growth of young producers is driven by demand-side accumulation rather than productivity gaps. What these have in common is their dynamics: they build up gradually over the life cycle, are buffeted by idiosyncratic shocks, and erode when not renewed (Benkard, 2000; Argote et al., 1990). *Organization capital* is the durable, documented asset a firm builds out of running its operations: procedures, structured data, trained models, executable workflow, customer records. It is bookable, it is far less exposed to the shocks

that buffet period-to-period performance, and it goes stale chiefly as the environment moves on.

Artificial intelligence, on this view, is a technology for converting operations into durable assets. It transcribes, structures, and operationalizes what a firm does, turning activity that would otherwise leave no persistent trace into a stock. We model it as an increase in the conversion rate between the two, and ask what that hypothesis implies. Beraja and Talamàs (2026) share this conceptualization, treating AI as an organizational learning technology whose aggregate value is governed by how much learning firms have left to do. We work at a finer resolution, splitting the accumulated advantage into a less codified component and a more codified one and letting AI act only on the conversion rate between them.

The exercise has three products. The first is a set of life-cycle predictions for firm size and asset composition. The second is a characterization of post-adoption dynamics, stated exactly in Proposition 3, which is what lets the model speak to our empirical finding that responses build over many years. The third is a cross-sectional test built on the organization capital ratio, developed in Section E.3 and shown in Section E.4 to govern the private value of adopting as well.

Terminology. Atkeson and Kehoe (2005) use "organization capital" for the pair (A, s), which in our notation is idiosyncratic productivity. We reserve the term for the codified stock B, in line with the measurement literature (Prescott and Visscher, 1980; Eisfeldt and Papanikolaou, 2013; Corrado et al., 2009), where organization capital denotes the durable, bookable organizational asset rather than idiosyncratic productivity. This is the object empirical proxies target. The relabeling is terminological; no equation of Atkeson and Kehoe (2005) is altered by it.

E.1. Model Setup

E.1.1. Households

Time is discrete, $t = 0, 1, 2, \ldots$. There is a unit continuum of households with preferences $\sum_t \beta^t \log c_t$. Each household contains one worker and one manager, each supplying one unit of labor inelastically. Households own the initial capital stock and the firms existing at date 0, face sequences of worker wages w_t, managerial wages w_{mt}, and intertemporal prices p_t, and choose consumption subject to the usual present-value budget constraint. This side of the economy is exactly as in Atkeson and Kehoe (2005).

E.1.2. Firms

A firm is characterized by (A, B, s): idiosyncratic productivity, organization capital, and age. Define *effective specific productivity*

$$\tilde{A} = A\, B^{\psi}, \qquad \psi \geq 0, \tag{E.1}$$

and let a firm operating with one manager, capital k, and labor l produce

$$y = z_t\, \tilde{A}^{\,1-n} F(k, l)^n, \qquad n \in (0, 1), \tag{E.2}$$

with F homogeneous of degree one, z_t economywide productivity, and n the span-of-control parameter. Setting $\psi = 0$ recovers Atkeson and Kehoe (2005).

Idiosyncratic productivity. Conditional on operating, $A_{s+1} = A_s e_s$ with $\log e_s \sim N(\mu_s, \sigma^2)$ and the Atkeson-Kehoe declining drift

$$\mu_s = \kappa_1 + \kappa_2 \left[\tfrac{S-s}{S}\right]^2 \mathbf{1}\{s \le S\}. \tag{E.3}$$

We take A as a reduced form and stay agnostic about its content: it may be technical efficiency, product appeal, an accumulated customer base (Foster et al., 2016), the firm's position relative to rivals, or the tacit, non-codified knowledge held by its workforce—the routines and judgment that have not been written down and that leave with the people who hold them. What (E.3) imposes is only the shape of the resulting life cycle—rapid gains when young, a peak, and erosion at rate $|\kappa_1|$ once mature.

Organization capital. Conditional on operating,

$$B_{s+1} = (1 - \delta_B) B_s + \chi A_s, \qquad B_0 = b_0 \tau_t, \tag{E.4}$$

where $\chi > 0$ is the rate at which operations are converted into codified form, $b_0 > 0$ is the codified content embodied in the entrant's blueprint (process templates, purchased systems), and δ_B is the depreciation rate of organization capital, reflecting obsolescence as documented processes, systems, and records go stale. A firm codifies out of what it does, so the flow into B is proportional to the scale of operations. Once codified, the asset is far less exposed to the innovation shocks: a bad quarter does not delete the process documentation, and a departing employee does not take the customer database with them.

The separation of the two stocks is motivated by a long tradition treating the binding constraint on organizational growth as the difficulty of transferring knowledge that resides in individuals and resists articulation (Polanyi, 1966; Nonaka, 1994). The constraint binds hardest early in the life cycle, before routines have stabilized and while learning depends on sparse and noisy feedback rather than on accumulated experience and documented practice (March, 1991; Grant, 1996). Knowledge held this way is also fragile: it erodes when not exercised (Argote et al., 1990) and degrades as practice is transmitted informally rather than written down. This is one reason organization capital accumulates as slowly as it does (Atkeson and Kehoe, 2005), and it is what χ represents in (E.4): the rate at which a firm's operations leave a trace the firm can retain and redeploy rather than remaining tied to the people who generated them.

Relation to other forms of intangible capital. Organization capital as modeled here shares the defining features of intangible capital more generally. First, Crouzet et al. (2022) emphasize that intangible capital must be stored on some medium, and that where it is stored governs who can use it and who can appropriate its returns. That observation is the organizing principle of our two-stock structure: idiosyncratic productivity is stored in people, arrangements, and relationships and is lost when they change, while organization capital is stored on durable media the firm controls and survives their turnover to a much greater degree. The parameter χ is the rate of transfer from the first toward the second,

and Assumption 1 is the claim that AI acts on that rate. Second, like all capital, the stock accumulates and depreciates, though δ_B here represents obsolescence of content rather than physical decay—documentation and records go stale because the environment they describe has moved on. Third, and more specifically, organization capital in (E.4) is a *by-product of production*: the flow χA_s is proportional to the scale of operations rather than freely chosen. This is the property that data has in Veldkamp and Chung (2024), where data is generated by economic activity rather than produced for its own sake, and it distinguishes organization capital sharply from R&D capital, which is a deliberate investment flow. A large codified stock is not available for purchase as such; it accumulates through operating and recording. Fourth, the stock is largely non-rival within the firm in the sense of Jones and Tonetti (2020): a documented procedure or a structured record can be used by every employee at once without depletion.[63]

Two features set organization capital apart, and both follow from firm specificity. First, it is not shared across firms. Data can be sold and patents licensed or disclosed (Liu et al., 2025), and both raise questions about the efficient allocation of a non-rival good across users; organization capital is non-rival within the firm but has little value outside it. Second, firm specificity also tends to make organization capital relatively excludable, which is worth noting since limited excludability is often treated as characteristic of intangibles (Crouzet et al., 2022). Patents rely on legal protection and trade secrets on secrecy, whereas a firm's documented routines are of limited use elsewhere, so specificity does some of the work that legal or contractual protection does for other intangibles. Two further consequences follow. Because it is largely neither traded nor separable from the firm, organization capital has no market price, which is why it must be inferred from firm behavior rather than read off transactions. And because little of it can be reallocated, it is largely destroyed at exit rather than sold, unlike physical capital or patents, which survive their owners.[64]

Firm specificity. Both stocks are firm-specific. Little spills over to entrants or rivals, and exit destroys (A, B) jointly, preserving the Atkeson-Kehoe assumption that the accumulated advantage of a closing producer "is lost permanently."

Define the *organization capital ratio* $r_s \equiv B_s / A_s$, with $r_{s+1} = [(1-\delta_B)r_s + \chi]/e_s$. Growth in effective productivity decomposes as

$$\Delta \log \tilde{A}_s = \underbrace{\Delta \log A_s}_{\text{idiosyncratic, } \mathbb{E}=\mu_s} + \psi \underbrace{\log\left[(1-\delta_B) + \chi / r_s\right]}_{\text{organization capital, } \equiv \gamma_s}. \tag{E.5}$$

[63] Non-rivalry generates increasing returns: scaling (A, B, k, l) by λ scales output by $\lambda^{1+\psi(1-n)}$. This is where (E.2) departs from the workhorse formulations in Crouzet and Eberly (2019) and Crouzet and Eberly (2021), in which intangibles enter as an additional rival input alongside physical capital; here B shifts $\tilde{A}$ outside $F(k, l)$, which is what converts non-rivalry into the scale economies above.

[64] While codified organization capital may be more portable (Lev and Radhakrishnan, 2005), the evidence suggests that codification alone is not enough: Zander and Kogut (1995) find that codifiability and teachability speed internal transfer but leave the hazard of imitation by competitors unaffected; Rivkin (2001) shows that a firm can replicate itself from an accurate template while rivals attempting to copy the same configuration fail; and Gibbons and Henderson (2012) argue that practices resting on relational contracts resist copying even where competitors understand them well.

The term γ_s refers to the growth rate of organization capital.

Capital and labor are freely mobile each period, so the input choice is static:

$$\max_{k,l} z_t \tilde{A}^{1-n} F(k,l)^n - r_t k - w_t l - w_{mt}.$$

Because F is homogeneous of degree one and there is a single fixed factor, all operating firms use inputs in the same proportions.

E.1.3. Entry, exit, and firm value

The decision to operate is dynamic. Firm value satisfies

$$V_t(A,B,s) = \max\left\{0,\ d_t(\tilde{A}) - w_{mt} + \tfrac{p_{t+1}}{p_t}\mathbb{E}_s V_{t+1}\big(Ae, (1-\delta_B)B + \chi A, s+1\big)\right\}. \tag{E.6}$$

The maximization in (E.6) is over the firm's only dynamic choice: the binary decision to operate this period or to exit. The first argument is the value of exit—the firm produces nothing and both stocks are destroyed permanently. The second is the value of operating: current variable profits $d_t(\tilde{A})$, which equal revenue net of payments to capital and labor at the static optimum, minus the fixed cost of hiring the manager, plus the discounted expected continuation value, where the expectation is over the innovation e to idiosyncratic productivity drawn from its age-s distribution.

Entry. Entry works as in Atkeson and Kehoe (2005). Before a new firm can enter in period t, a manager must spend period $t-1$ preparing and adopting a blueprint for constructing the firm, and the blueprint determines the firm's initial state. Blueprints adopted in period $t-1$ embody the frontier of knowledge regarding the design of firms at that point in time; this frontier τ_t evolves exogenously, growing at rate g_τ. Thus a firm built in $t-1$ starts period t with idiosyncratic productivity τ_t and organization capital $b_0\tau_t$, so its state is $(A,B,s) = (\tau_t,\, b_0\tau_t,\, 0)$: the blueprint embodies not only an initial market position but also a starter stock of codified material—process templates, purchased software and systems—scaled by the same frontier. Free entry into the activity of preparing blueprints drives the net value of entry to zero, which pins down the managerial wage:

$$w_{mt} = \tfrac{1}{1+i}\, V_{t+1}\big(\tau_{t+1},\, b_0\tau_{t+1},\, 0\big), \tag{E.7}$$

where i is the equilibrium interest rate.

Managers. Managers are the economy's fixed factor and are used in two activities. Let f_t denote the measure of managers spending period t preparing blueprints for firms that will enter in $t+1$, and let M_t denote the measure of firms operating in t, each of which employs exactly one manager as its fixed factor. Since each household supplies one manager, the resource constraint for managers is $f_t + M_t = 1$.

As in Atkeson and Kehoe (2005), two features of (E.6) matter below. V is convex in A—the flow is linear in $\tilde{A}$ and the continuation is truncated at zero—and young firms typically have $d < w_m$: organization rents are back-loaded and young firms operate on option value.

E.1.4. Balanced growth and stationary composition

Following Atkeson and Kehoe (2005), we define a balanced growth path as an equilibrium in which the quality of the best available blueprint τ_t and economywide productivity z_t grow at constant rates g_τ and g_z; aggregate output, consumption, capital, and the two wages grow at a common constant rate; the interest rate, the entry measure f_t, and the measure of operating firms M_t are constant; and the distribution of firm states is stationary once each cohort is detrended by its blueprint—$(A_s/\tau, B_s/\tau)$ has the same law for every cohort, so each generation of firms traverses the same life cycle, scaled up by the frontier. Along such a path effective productivity across cohorts grows at rate $(1+g_\tau)^{1+\psi}-1$, since both stocks scale with the cohort blueprint, and the aggregate growth rate follows from the Atkeson-Kehoe formula with this embodied rate.

Stationarity of the firm's asset *composition* requires a restriction linking the two decay rates.

Proposition 1 (Stationary composition)**.** *Consider a mature firm drawing from the terminal distribution with drift κ_1. Along the median path the organization capital ratio converges to*

$$r^* = \frac{\chi}{e^{\kappa_1} - (1-\delta_B)}$$

if and only if $\delta_B > 1 - e^{\kappa_1}$, and the fixed point is globally stable on $\mathbb{R}_+$. If $\delta_B \le 1 - e^{\kappa_1}$, then $r_s \to \infty$ from any initial condition.

Proof. On the median path $e_s = e^{\kappa_1}$, so the organization capital ratio follows the affine map

$$r_{s+1} = \phi(r_s) \equiv a\, r_s + c, \qquad a = (1-\delta_B)e^{-\kappa_1}, \quad c = \chi e^{-\kappa_1} > 0.$$

Iterating, $r_s = a^s r_0 + c\,(1-a^s)/(1-a)$ for $a \neq 1$. If $a < 1$—that is, $1-\delta_B < e^{\kappa_1}$, or $\delta_B > 1 - e^{\kappa_1}$—then $a^s \to 0$ and $r_s \to c/(1-a) = \chi/(e^{\kappa_1} - (1-\delta_B)) = r^* > 0$ monotonically from any $r_0 \ge 0$, and since $|\phi'| = a < 1$ the fixed point is a global contraction. If $a \ge 1$, then $a^s r_0$ is nondecreasing and the geometric sum $c(1 + a + \cdots + a^{s-1})$ diverges, so $r_s \to \infty$. □

The economics of the condition is that organization capital must go stale faster than the firm's idiosyncratic productivity advantage erodes; otherwise the documentation outlives the business it documents. At the Atkeson-Kehoe drift, the requirement is $\delta_B > 1 - e^{\kappa_1} = 10.8\%$ per year, below the 15–20% depreciation rates for organization capital commonly used in the literature (Eisfeldt and Papanikolaou, 2013; Crouzet et al., 2022).[65]

E.1.5. Calibration and the baseline life cycle

Macro parameters are from Atkeson and Kehoe (2005): $\beta = .98$, $n = .85$, physical capital share 19.9%, $i = 4.1\%$, $g = 2.02\%$. The parameters of (E.3) are their estimates from

[65] Direct evidence suggests this is comfortably satisfied: Benkard (2000) finds that production experience in aircraft manufacturing depreciates substantially over time, Argote et al. (1990) estimate very rapid depreciation of the accumulated experience stock in shipbuilding, and Foster et al. (2016) estimate an annual depreciation rate of roughly 36% for the accumulated demand stock. Abis and Veldkamp (2024) estimate a monthly depreciation rate of 3% for data, implying an annual rate of 31%. Accumulated firm advantages go stale quickly; the stationarity condition is therefore a description of the data, not a strained requirement.

the 1988 LRD panel (Davis et al., 1996). The depreciation rate δ_B is set to 0.15, the value conventionally used in the organization capital literature. The codification-side parameters are inferred based on our estimates and discussed in Section E.5.

Table E.1. Parameters used in the numerical illustrations

	Parameter	Value	Source
Idiosyncratic productivity	κ_1	-0.1139	Atkeson and Kehoe (2005)
	κ_2	0.1741	Atkeson and Kehoe (2005)
	σ	0.1945	Atkeson and Kehoe (2005)
	S	100	Atkeson and Kehoe (2005)
Organization capital	ψ	0.50	based on reduced-form estimates (Section E.5)
	χ	0.15	illustrative
	δ_B	0.15	$> 1 - e^{\kappa_1} = 0.108$; cf. depreciation estimates
	b_0	0.5	entrants start half-codified
AI treatment intensity	χ'/χ	2.0	based on reduced-form estimates (Section E.5)
Adoption values	ρ	0.85	discount $\times$ survival

All figures plot median paths (e_s at its median), isolating life-cycle profiles from the dispersion generated by σ. Figure E.1 shows the baseline. Idiosyncratic productivity peaks at 0.59 log points around age 20 as the drift crosses zero; effective productivity—firm size—peaks at 1.18 log points at age 22, growing at 13.0% per year at entry (6.0 points from idiosyncratic productivity plus $\psi\gamma_0 = 7.0$ points from codification). Organization capital is a lagged, smoothed integral of idiosyncratic productivity: its growth rate γ_s starts at 14.0% per year, declines monotonically, and turns negative at age 26—exactly the age at which the organization capital ratio crosses χ/δ_B—before converging to κ_1 as the composition stabilizes at $r^* = 3.54$.

E.2. AI as a Codification Technology

Assumption 1 (Codification)**.** *AI raises the conversion rate: $\chi \mapsto \chi' > \chi$ from the date of adoption onward. The idiosyncratic productivity process $\{\mu_s, \sigma\}$, the depreciation rate δ_B, and the blueprint content b_0 are unchanged.*

Modeling AI as an increase in χ reflects what these systems do when trained on a firm's own records of work. The raw material is the record the firm's operations generate: not only the accumulated output of individual employees—code repositories, service transcripts, clinical documentation—but also the machine-generated trace of the operation itself, such as line speeds, cycle times, and quality readings, all of which previously served only to document what had been done. Training a model on it converts that record into an operational asset, and does so far faster than the alternative: what would otherwise diffuse slowly through mentoring, documentation, and turnover is extracted in a single pass, so the problem-solving approaches and conventions embedded in the work of particular employees become available to everyone else in the organization almost at once

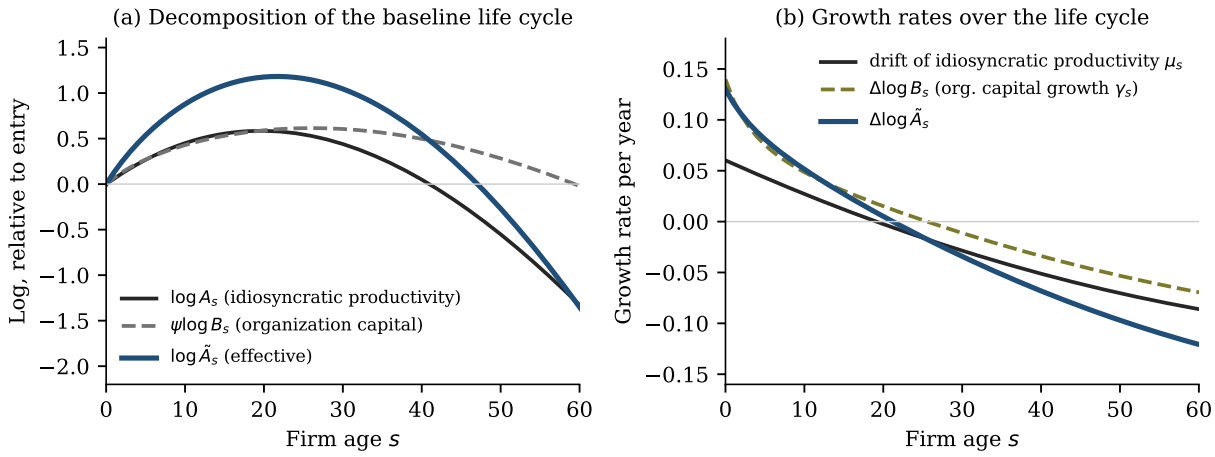


Figure E.1. The baseline (pre-AI) firm life cycle. Panel (a): decomposition of effective productivity into idiosyncratic productivity and the contribution of organization capital. Panel (b): the drift of idiosyncratic productivity μ_s, organization-capital growth γ_s, and effective productivity growth. Median paths at the parameters of Table E.1.

(Brynjolfsson et al., 2025b; Singhal et al., 2023).[66]

This is precisely the transformation (E.4) describes, and Appendix B.3 documents it across eight firms: the flow into organization capital is proportional to the scale of operations because operations are what the model is trained on, and χ governs how much of that activity is converted into a form the firm retains. Cullen et al. (2026) make the point explicit—expertise that had been an attribute of individual workers becomes, once embedded in a model trained on their output, an asset the firm owns and can deploy without depending on any one of them. Where the record is machine-generated rather than human, the conversion is the same but the constraint it relaxes is different: a firm cannot run experiments on a production line without stopping it, and a model of the line, built from the line's own history, removes that limit. Another complementary channel is that AI extracts more from a given operational record than earlier methods could (Agrawal et al., 2022), so a firm need not wait to accumulate decades of history before its records support a usable asset. Moderna, discussed in Appendix B.3, illustrates the point. Without a commercial product until 2020, it built predictive models from thousands of development experiment data points that let its process engineers avoid failures they had not yet encountered.

AI relative to earlier codification technologies. The mechanism in (E.4) is not new to AI. Enterprise resource planning, customer relationship management, and workflow systems are all technologies that convert operating practice into a durable, firm-held record, and

[66] For example, Brynjolfsson et al. (2025b) show that generative AI systems capture precisely the behaviors hardest to write into explicit procedure—the judgment and communication patterns of a firm's strongest performers—and diffuse them through real-time guidance. That is a change in how much of a firm's existing capability becomes shareable, not a change in the capability itself.

the productivity literature has long treated their returns as contingent on complementary organizational investment rather than on the software itself (Brynjolfsson and Hitt, 2000; Bresnahan et al., 2002). What differs is the range of practice that can be converted. Earlier systems were bounded by what Autor (2015) calls Polanyi's paradox: a task could be encoded only insofar as it could be explicitly specified, and much of what firms know how to do had never been written down—not because it was unwritable, but because articulating it was costly and no one had paid that cost. AI relaxes this constraint, because it infers regularities from demonstration rather than requiring them to be stated in advance, and can therefore capture practice that its own practitioners cannot fully articulate. What was previously a barrier to codification becomes, at least partly, an input to it. In the model this is an increase in χ rather than a new mechanism, which is what makes the framework applicable to AI and to earlier codification technologies alike; it also implies that the firms with most to gain are those whose operations were hardest to write down, which the case studies in Appendix B.3 illustrate.

Proposition 2 (Properties of codification)**.** *Under Assumption 1: (i)* Permanent level effect. *B_s is affine in χ with a fixed anchor $b_0\tau$, and effective productivity is permanently higher by $\psi \log(\chi'/\chi)$. (ii)* Composition shift. *The stationary organization capital ratio rises to $(\chi'/\chi)\, r^*$: $\log r_s$ is permanently higher by $\log(\chi'/\chi)$. (iii)* Idiosyncratic productivity process unchanged. *The drift profile and innovation variance of A are unaffected.*

Proof. Solving (E.4) forward,

$$B_s = \underbrace{(1-\delta_B)^s\, b_0\tau}_{\text{depreciated endowment}} + \chi \underbrace{\sum_{j=0}^{s-1}(1-\delta_B)^{s-1-j}A_j}_{\text{depreciated sum of past operations}}, \tag{E.8}$$

so organization capital is the sum of two pieces. The first is what remains of the codified endowment the firm entered with; it does not involve χ at all. The second is the depreciated sum of everything the firm has converted since birth; it is exactly proportional to χ. "Affine with a fixed anchor" means just this: doubling the conversion rate doubles the accumulated piece while leaving the endowment piece untouched. Early in life the endowment still matters, so the effect of a higher χ is partial; as the endowment depreciates away, organization capital becomes proportional to χ.

(i) Equation (E.8) holds for any shock path. An adopter and its non-adopting counterfactual share the same path of A (by part (iii)) and the same anchor term, so

$$\frac{B_s^1}{B_s^0} = \frac{(1-\delta_B)^s b_0\tau + \chi'\Sigma_s}{(1-\delta_B)^s b_0\tau + \chi\,\Sigma_s} \longrightarrow \frac{\chi'}{\chi} \quad \text{as } s \to \infty,$$

where Σ_s is the depreciated sum in (E.8), since the anchor terms vanish relative to Σ_s. The log gap in organization capital converges to $\log(\chi'/\chi)$, and by (E.1) the log gap in effective productivity converges to $\psi \log(\chi'/\chi)$. (ii) Immediate from Proposition 1: r^* is linear in χ. (iii) By inspection: the law of motion (E.3) of A involves neither χ nor B. □

Figure E.2 illustrates. Adoption at entry raises growth at birth from 13.0% to 24.6%

per year, raises the peak from 1.18 to 1.53 log points at the same age, and shifts the entire mature path up by the permanent $\psi \log(\chi'/\chi) = 0.347$: the gap reaches 0.325 by age 10 and is essentially complete by age 20. Under an illustrative exit threshold the lifespan extends from 53 to 56 years; more generally, because B is immune to the innovation shocks, adopters carry a larger buffer against bad draws, mature exit falls, and exits follow prolonged decline rather than single shocks. Panel (b) shows the composition ratio converging to a permanently higher plateau, $\log r^*$ shifting up by exactly $\log(\chi'/\chi)$.

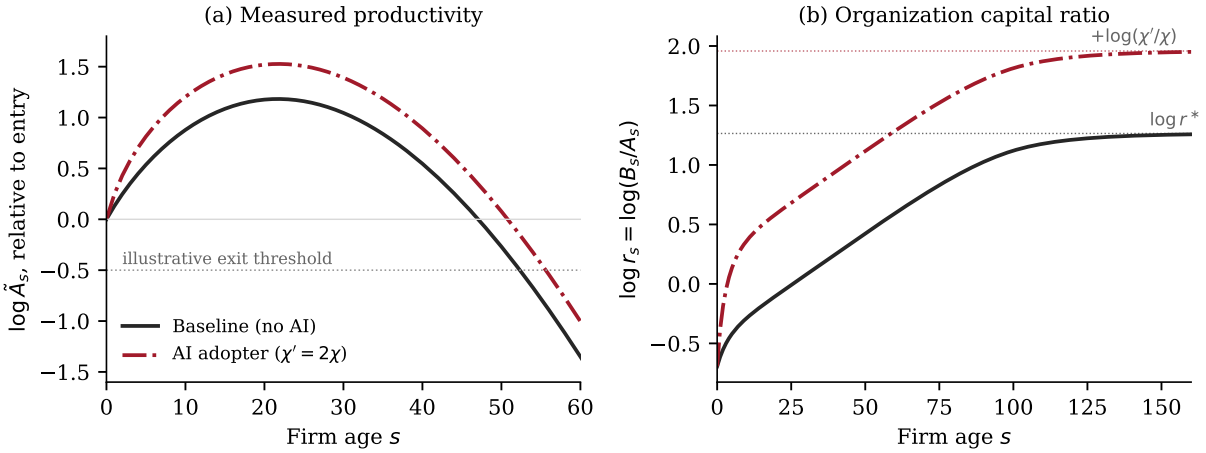


Figure E.2. Life-cycle paths under no AI and under adoption at entry. Panel (a): measured productivity $\log \tilde{A}$; the dotted line is an illustrative exit threshold, crossed at ages 53 and 56. Panel (b): the organization capital ratio $\log r_s$, converging to $\log r^*$ in the baseline and to $\log r^* + \log(\chi'/\chi)$ under AI.

Two further predictions follow directly from (E.5). *Persistence:* post-adoption, a larger share of effective productivity is the smoothed integral of past idiosyncratic productivity, so a shock to A passes through with long-run elasticity $1 + \psi$ at a rate governed by δ_B, and the autocorrelation of size growth rises. *Selection into adoption:* conditional on age and size, the firms with the most to gain are those with low r—large operations that have codified little of what they do—a point developed in Sections E.3 and E.4.

E.3. Event-Study Dynamics

The bridge from the model to panel evidence is that the post-adoption gap in organization capital obeys an exact first-order recursion, with coefficients built from observables.

Proposition 3 (Event-study dynamics)**.** *Consider a firm adopting at age s_0 and its never-adopting counterfactual with the same shock path. Let B^1, B^0 denote the two organization-capital paths, $x_h \equiv (B^1_{s_0+h} - B^0_{s_0+h})/B^0_{s_0+h}$ the proportional gap, and $G_h \equiv B^0_{s_0+h+1}/B^0_{s_0+h} = e^{\gamma_{s_0+h}}$ the counterfactual growth factor of organization capital. Then for all $h \geq 0$:*

(i) Exact AR(1). $x_0 = 0$ *and*

$$x_{h+1} = \rho_h\, x_h + (1 - \rho_h)\,\frac{\chi' - \chi}{\chi}, \qquad \rho_h = \frac{1 - \delta_B}{G_h} = (1 - \delta_B)e^{-\gamma_{s_0+h}}.$$

(ii) Common plateau. $x_h \to (\chi' - \chi)/\chi$ *regardless of* s_0 *and of the realized path of* A*: the log gap in organization capital converges to* $\log(\chi'/\chi)$ *for every adoption age.*

(iii) Proportionality at every horizon. *The measured-productivity gap is exactly* ψ *times the organization-capital gap:* $\log \tilde{A}^1_{s_0+h} - \log \tilde{A}^0_{s_0+h} = \psi \log(1 + x_h)$ *for all* h*.*

(iv) Impact response. *Setting* $h = 0$*, the first-period response of organization-capital growth is*

$$\Delta\gamma(s_0) = \log \frac{(1 - \delta_B) + \chi'/r_{s_0}}{(1 - \delta_B) + \chi/r_{s_0}},$$

strictly decreasing in the pre-adoption organization capital ratio r_{s_0}*.*

Proof. Let $D_h = B^1 - B^0$ at horizon h. Subtracting the two accumulation equations, $D_{h+1} = (1 - \delta_B)D_h + (\chi' - \chi)A_{s_0+h}$, with $D_0 = 0$. Dividing by $B^0_{s_0+h+1} = G_h B^0_{s_0+h} = (1 - \delta_B)B^0_{s_0+h} + \chi A_{s_0+h}$ and using $\chi A_{s_0+h}/B^0_{s_0+h} = G_h - (1 - \delta_B)$ gives (i). Since $\rho_h \in (0, 1)$, x_h converges to the fixed point $(\chi' - \chi)/\chi$, which does not involve the A path, giving (ii). Part (iii) is immediate from $\log \tilde{A} = \log A + \psi \log B$ and identical A paths under Assumption 1(iii). For (iv), $x_1 = (1 - \rho_0)(\chi' - \chi)/\chi$ with $\rho_0 = (1 - \delta_B)/[(1 - \delta_B) + \chi/r_{s_0}]$; the first-period growth-rate difference is $\log(B^1_{s_0+1}/B^1_{s_0}) - \log(B^0_{s_0+1}/B^0_{s_0}) = \log(1 + x_1)$, and substituting yields $1 + x_1 = [(1 - \delta_B) + \chi'/r_{s_0}]/[(1 - \delta_B) + \chi/r_{s_0}]$. □

Figure E.3 displays the gap dynamics for adoption at ages 5, 30, and 60: a common plateau reached at different speeds, with the productivity panel exactly ψ times the organization-capital panel at every horizon. Idiosyncratic productivity A is *identical* for adopters and counterfactuals by Assumption 1, so all the divergence in both panels is due to organization capital.

E.3.1. Why adjustment is slow

The convergence speed in Proposition 3(i) is $1 - \rho_h = 1 - (1 - \delta_B)e^{-\gamma_s}$, approximately $\delta_B + \gamma_s$: the depreciation rate of organization capital *plus the growth rate of the untreated stock*. The intuition is a savings account. Each year the firm deposits χA and the balance depreciates at δ_B; AI raises the deposit rate. How fast the balance reprices to the new deposit rate depends on how large deposits are relative to the accumulated balance—and that ratio is $\gamma_s + \delta_B$. A young firm's stock is mostly recent deposits, so it reprices in a few years. A mature firm's stock is mostly legacy accumulation, and for such a firm γ_s is near zero or negative, so repricing is slow: at the calibration, half-lives are 3, 4, and 6 years for adoption at ages 5, 30, and 60, and the mature limit ($\gamma_s \to \kappa_1$) is 14.6 years.

The magnitudes matter for how we read the evidence. In our empirical work, organization capital and measured productivity take many years to respond fully to AI investment. In this model that is not evidence of adjustment costs, implementation lags, or slow organizational learning—none of which are present. It is what the accounting of a stock-flow system delivers for mature firms: for any firm whose organization capital is stable or declining ($\gamma_s \leq 0$), every depreciation rate below roughly 13–22% implies a half-life above five years, and the depreciation estimates cited in Section E.1 sit squarely in that range.

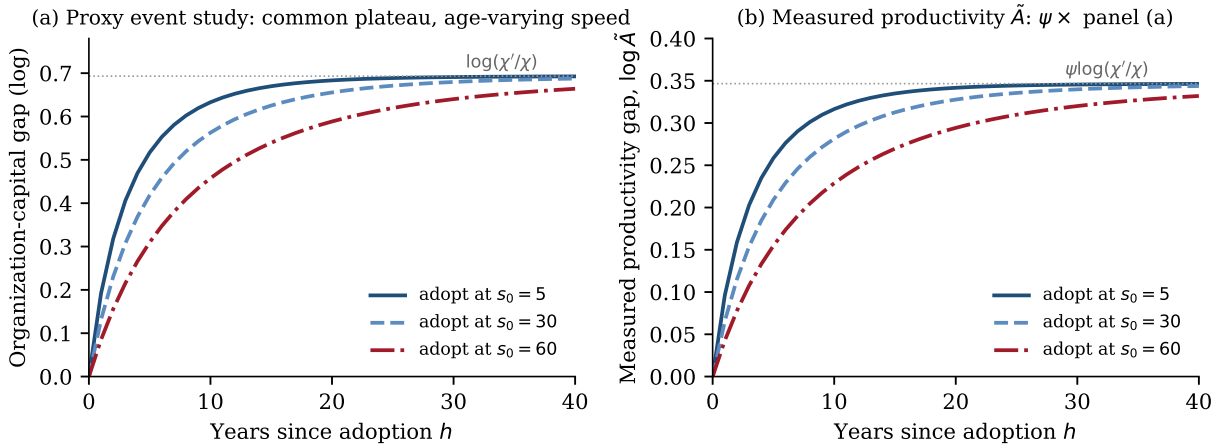


Figure E.3. Event studies around AI adoption at ages 5, 30, and 60, relative to never-adopting firms of the same age. Panel (a): the organization-capital (proxy) gap converges to the common plateau $\log(\chi'/\chi)$ at speeds $\delta_B + \gamma_s$ that fall with the pre-adoption organization capital ratio. Panel (b): the measured-productivity gap $\log \tilde{A}$ is exactly ψ times panel (a) at every horizon.

E.3.2. Heterogeneity test

Proposition 3 has a sharp cross-sectional implication. Both the impact response (part iv) and the convergence speed (part i, through γ_s) depend on the adopting firm's state only through its pre-adoption organization capital ratio r_{s_0}. Two firms with the same r respond identically *regardless of age*; the plateau, meanwhile, is common to all. It follows that at any finite horizon the treatment effect—on organization capital and on measured productivity—is *decreasing in the initial organization capital ratio*: low-r firms respond more on impact and converge faster toward the same eventual gain. The reason r_s is a sufficient statistic for the response is visible in the definition of γ_s itself: from (E.5) the untreated growth rate of organization capital is $\gamma_s = \log\big[(1-\delta_B) + \chi/r_s\big]$, so both objects that govern the response—the impact effect $\Delta\gamma(s) = \log \frac{(1-\delta_B)+\chi'/r_s}{(1-\delta_B)+\chi/r_s}$ and the convergence factor $\rho_h = (1-\delta_B)e^{-\gamma_s} = (1-\delta_B)\big/\big[(1-\delta_B) + \chi/r_s\big]$—depend on the firm's state through r_s and nothing else.

This is the heterogeneity test we take to the data: interact AI investment with the pre-adoption ratio of the organization-capital proxy to firm size, and predict larger medium-run responses for firms with low initial ratios, with the heterogeneity shrinking at long horizons as plateaus equalize.

Why not simply test for larger effects among young firms, since r_s rises with age in the model? Three reasons, one practical and two conceptual. The practical one: firm age is essentially unobserved in Compustat—incorporation and listing dates measure age at listing, not the age of the underlying business, and are especially unreliable for firms formed through mergers and spinoffs. The first conceptual reason is sufficiency: age matters only through the organization capital ratio, and idiosyncratic shocks disperse that ratio within a cohort, so age is a noisy proxy for the state variable that actually governs

the response—while the ratio itself is directly observable as organization capital relative to size. The second conceptual reason is confounding through the treatment intensity. Outside the strict model, χ'/χ need not be a constant: the productivity of AI plausibly *rises* with a firm's accumulated data and organizational experience, since models trained on richer operational histories are better models. If so, age-based heterogeneity confounds two opposing forces—older firms have lower capture speed (smaller effect) but possibly higher treatment intensity (larger effect)—with an ambiguous net sign, and conditioning on r isolates the capture mechanism.[67]

E.4. Endogenous Adoption over the Life Cycle

So far adoption has been exogenous. This section makes it a choice. A firm of age s_0 may pay a one-time irreversible cost κ, in units of entry-size output, to raise its conversion rate permanently from χ to χ'. Writing $\tilde{A}^1$ and $\tilde{A}^0$ for the adopting and non-adopting paths from s_0 onward and ρ for the product of the discount factor and the survival rate, the gross gain from adopting at s_0 is

$$G(s_0) \;=\; \sum_{h\geq 0} \rho^h \big[\tilde{A}^1_{s_0+h} - \tilde{A}^0_{s_0+h}\big], \tag{E.9}$$

which is proportional to the discounted gain in output and variable profits. A firm facing a one-time cost κ adopts at the age maximizing $\rho^{s_0}[G(s_0) - \kappa]$; we focus here on the shape of G, which is what governs both the adoption margin and the response heterogeneity.

Proposition 4 (Value of adopting)**.** *To first order in the treatment intensity,*

$$G(s_0) \;=\; \underbrace{\psi}_{\text{elasticity}} \times \underbrace{\bar{x}}_{\text{intensity}} \times \underbrace{W(s_0)}_{\text{remaining scale}} \times \underbrace{\Lambda(s_0)}_{\text{capture fraction}} ,$$

where $\bar{x} = (\chi' - \chi)/\chi$, $W(s_0) = \sum_{h\geq 0} \rho^h \tilde{A}^0_{s_0+h}$ *is the firm's discounted remaining size, and*

$$\Lambda(s_0) = \frac{\sum_h \rho^h \tilde{A}^0_{s_0+h}\, x_h(s_0)}{\bar{x} \sum_h \rho^h \tilde{A}^0_{s_0+h}} \in (0,1)$$

is the size-weighted fraction of the eventual plateau that the firm captures. Λ *is decreasing in the pre-adoption organization capital ratio* r_{s_0} *and hence, along the life cycle, decreasing in age;* W *is hump-shaped. The product is hump-shaped in* s_0*.*

The decomposition separates two economically distinct forces, plotted in Figure E.4. *Remaining scale* rises through the first two decades, because a young firm's discounted future size is dominated by a size peak it has not yet reached, and falls thereafter; at the calibration it peaks at age 17. *Capture fraction* falls monotonically—from 0.59 at entry to

[67] This confound signs the bias in our favor: if intensity rises with the accumulated stock, then low-r firms receive *weaker* treatments, which works against finding larger responses for them. The organization-capital-ratio test is therefore conservative with respect to the leading violation of the constant-intensity assumption.

0.45 at age 12 and 0.27 at age 40—because a firm that has already codified much of its operation converges to the plateau slowly, by the same ρ_h of Proposition 3 that governs Section E.3.1: what makes mature adopters' event studies slow makes their adoption less valuable. The gain peaks at age 12, before both the scale peak at 17 and the firm's size peak at 22. The hump is, if anything, conservative about how late in life adoption is worth most.

Note that the state variable doing the work is again the organization capital ratio, not age: at age 15, a firm holding half the modal organization capital ($r = 0.41$) has a gain of 4.65 entry-size units against 4.30 for the modal firm and 3.85 for one holding twice the modal stock—a 21% spread across a factor of four in organization capital intensity, entirely through Λ. The adoption margin and the treatment-response heterogeneity of Section E.3.2 are the same economics: AI is worth most, and works fastest, in firms that are large relative to what they have codified. One might have expected the return to codification to be largest for the largest firms, since the flow gain scales with size; it is not, because the largest firms are the oldest and have already converted the bulk of what they do into durable form.

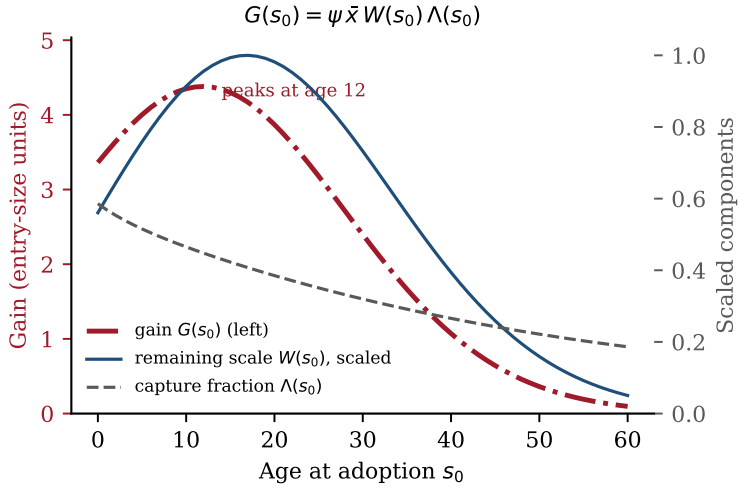


Figure E.4. The gross gain from adopting at age s_0 (left axis) and its two components from Proposition 4: discounted remaining scale $W(s_0)$ and capture fraction $\Lambda(s_0)$ (right axis, scaled).

This age profile is visible in the data. Figure E.5 plots the share of a firm's organization-capital-related AI jobs against firm age, averaged within age bins across the 2024 cross-section. The pattern is hump-shaped in the way the model implies. Very young firms invest relatively little—about 0.5% of the workforce at ages zero to four—despite having the most left to codify. Investment then rises steeply, roughly two and a half times, to peak near 1.2% between ages five and thirteen, and declines steadily thereafter, falling to around 0.3% among firms over seventy. Two forces of Proposition 4 map onto the two arms of the hump: the rise reflects growth in remaining scale, as young firms have not yet reached the size over which the return to codification is earned, while the decline reflects the falling capture fraction, as older firms have already converted much of what they do into durable form and their existing stock dilutes what they add.

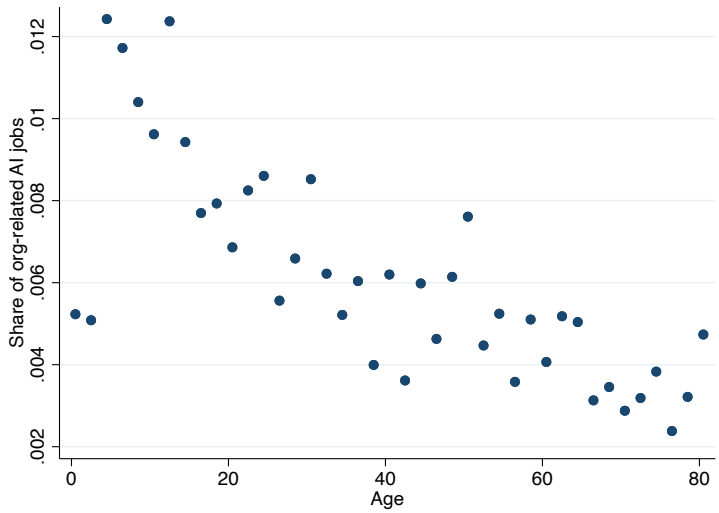


Figure E.5. Share of organization-capital-building AI jobs, by firm age. Firm age is measured in 2024 using founding year from Revelio. Each point is the average share across firms within a two-year age bin, with firm age top-coded at 80.

E.5. Taking the Model to Data

We conclude with the intuition for how the model organizes panel evidence on AI investment, organization capital, and productivity. A full estimation is beyond this paper; here we record only how the reduced-form objects map into the model's parameters.

Empirically, we measure organization capital by capitalizing SG&A, and read the resulting stock as a constant multiple of B; the unknown scale never matters, because every relevant statistic is a growth rate, a gap, or a ratio in which it differences out. The obvious concern with this proxy is that SG&A bundles spending that builds the codified stock (systems, documentation, data infrastructure) with spending that builds idiosyncratic productivity (customer acquisition, and knowledge that remains uncodified), so the constructed measure is better read as B contaminated by some component of A. This matters less than it might appear, because our estimates come from *changes* around adoption rather than from levels. Under Assumption 1 AI leaves A untouched, so the contaminating component does not respond to adoption: it shifts the level of the proxy without contributing to the event-study response, and the response still traces the organization-capital gap g_h that Proposition 3 describes.[68]

Recovering ψ and χ'/χ

Our specification is not an event study but a long difference: we regress the change in our proxy for organization capital, and separately the change in measured productivity, on the change in AI investment over the sample window. Proposition 3 maps directly into this design. Let H denote the length of the window and $g_H \equiv \log(1 + x_H)$ the log

[68] Because the contamination inflates the measured stock without inflating its response, it attenuates the estimated growth rates and the implied plateau, so $\widehat{\chi'/\chi}$ is biased toward zero—a conservative direction.

organization-capital gap that adoption has opened up by the end of it. The first regression recovers g_H and, by Proposition 3(iii), the second recovers ψ times the same object:

$$\beta_B = g_H, \qquad \beta_{\tilde{A}} = \psi\, g_H. \tag{E.10}$$

The elasticity is therefore the ratio of the two coefficients, and the treatment intensity is obtained by scaling the first up by however far convergence has run:

$$\psi = \frac{\beta_{\tilde{A}}}{\beta_B}, \qquad \log\frac{\chi'}{\chi} = \frac{g_H}{\varphi_H}, \qquad \varphi_H \equiv \frac{x_H}{\bar{x}} = 1 - \prod_{h<H} \rho_h. \tag{E.11}$$

The two objects are identified on very different terms, and it is worth being explicit about which is robust. The ratio in (E.11) requires no assumption about the horizon: by Proposition 3(iii) the productivity gap is ψ times the organization-capital gap at *every* h, so a partially converged long difference recovers ψ as cleanly as a fully converged one. It is also invariant to how the AI investment measure is scaled, since the normalization enters both coefficients and cancels, and it is robust to anything that moves B for reasons outside the model, since such forces move measured productivity by ψ times as much. Recovering χ'/χ, by contrast, requires taking a stand on φ_H and on what a unit of the treatment represents.

In our data the instrumented long-differences coefficient is 0.24–0.30 for organization capital (columns 3 and 6 of Table C.30) and 0.11–0.14 for firm productivity (columns 3 and 6 of Table 4), so (E.11) gives $\psi \approx 0.5$. Taking the view that our sample window captures roughly half of the eventual response—plausible given the half-lives implied in Section E.3.1—sets $\varphi_H \approx 0.5$ and hence $\log(\chi'/\chi) \approx 0.6$, or $\chi'/\chi \approx 2$: AI roughly doubles the rate at which operations are converted into organization capital. These are the values reported in Table E.1.

Both numbers should be read as illustrative magnitudes rather than precise estimates, because capitalized SG&A is a noisy measure of B and its errors do not obviously wash out. Two biases work in opposite directions. To the extent that the proxy contains a component of idiosyncratic productivity that does not respond to adoption, the measured stock is inflated without its response being inflated, attenuating β_B and thus overstating ψ while understating the plateau. To the extent that AI expenditure is itself partially booked in SG&A, part of β_B is accounting mechanics rather than organization-capital accumulation, which pushes ψ the other way. We report $\psi \approx 0.5$ and $\chi'/\chi \approx 2$ as the central case without claiming the two biases cancel.

E.6. Extension: OC as a Substitute for Measured Inputs

The environment above treats organization capital as a productivity shifter: it raises the marginal product of capital and labor, so a firm that accumulates B optimally expands its purchased inputs alongside its output. That is not what we find: sales respond to AI investment while employment and PP&E do not, so the coefficients on sales and on sales per worker are close to equal. This subsection sketches an extension in which that is the case. Organization capital is intangible: capitalized SG&A is by construction not part of PP&E, and the employees who build it are already counted in headcount. If the codified

stock substitutes for purchased inputs rather than merely raising their productivity—a digital twin replacing physical trial capacity, a knowledge base replacing staffed support, route optimization replacing trucks and miles—then the firm's effective input bundle can expand while its measured inputs do not. Write

$$y = z\, A^{1-n}\, B^{\beta}\, \big[\underbrace{X + \lambda B}_{\tilde{X}}\big]^{n}, \qquad X = k^{\alpha} l^{1-\alpha}, \tag{E.12}$$

where X is purchased at market prices and λB is supplied by the firm's own codified stock at zero marginal cost. Organization capital enters twice, and both terms are necessary: β governs whether output rises, and λ whether purchased inputs rise or fall. With $\beta = 0$, the firm holds output fixed and simply swaps purchased inputs for its own stock, so codification would be a pure cost saving, predicting flat sales and *falling* capital and employment, which is inconsistent with our empirical findings. Setting $\lambda = 0$ recovers the baseline environment exactly, so (E.12) adds one parameter rather than replacing the model.

Solving the static problem gives $\tilde{X} \propto A\, B^{\beta/(1-n)}$ and hence

$$y \;\propto\; A\, B^{\beta/(1-n)}, \tag{E.13}$$

so the extension is *observationally equivalent to the baseline for output*: the composite $\beta/(1-n)$ plays exactly the role ψ plays there, and output data cannot separate the two. We therefore write the extension in terms of the same ψ, with $\beta = \psi(1-n) = 0.075$ its structural counterpart given n. The life-cycle profiles in Figures E.1 and E.2 are unchanged, as are the event-study dynamics of Proposition 3, which concern the law of motion (E.4) rather than the production function. What the extension changes is the input side. Letting $s \equiv \lambda B/\tilde{X}$ denote the share of the effective bundle supplied by organization capital,

$$\frac{d\log X}{d\log B} = \frac{\psi - s}{1-s}, \qquad \frac{d\log(y/l)}{d\log B} = \psi - \frac{\psi - s}{1-s}. \tag{E.14}$$

Figure E.6 plots these. The baseline is the case $s = 0$: purchased inputs move with output and sales per worker does not respond. As s rises, a growing part of the expansion is met from the firm's own stock, purchased inputs respond less, and sales per worker responds more. Flat measured inputs correspond to the knife-edge $s = \psi$, at which purchased inputs are exactly unchanged and sales per worker moves one for one with sales—the pattern in our estimates.[69] Panel (b) shows that $s \propto B/y$, which is the observable proxy-to-size ratio of Section E.3.2; it is roughly flat over the first three decades of a firm's life and rises thereafter, and it shifts up permanently for an adopter.

To summarize, modeling AI as a codification technology—a change in the rate at which a firm's operations become durable, documented assets—yields a permanent level effect on firm size that arrives gradually, a permanent shift in the composition of firm assets, and

[69] In a more general specification that replaces $X + \lambda B$ with a CES aggregator of X and λB—which nests (E.12) as the limiting case of an infinite elasticity of substitution—flat inputs can arise for a range of substitution elasticities and input shares, rather than at a single value of s.

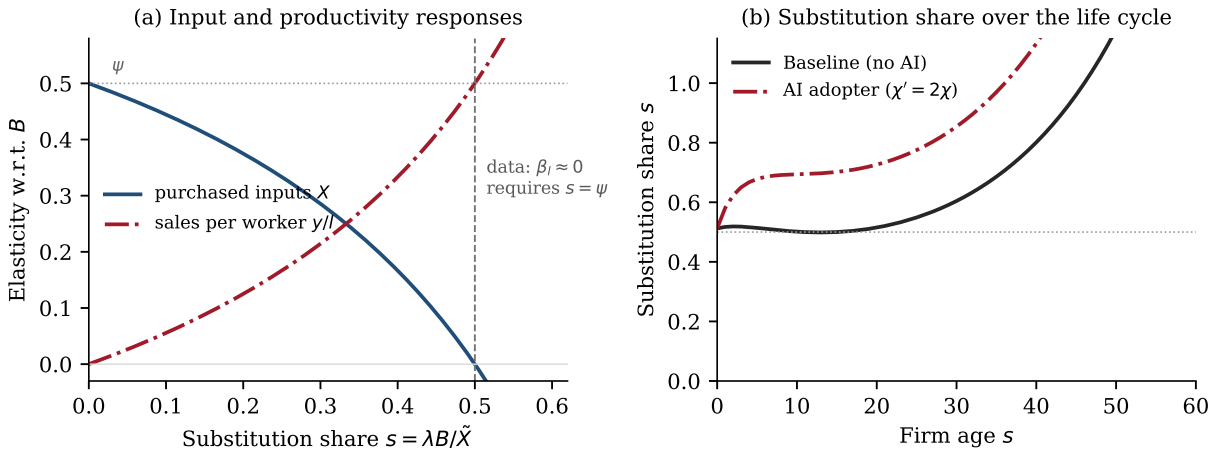


Figure E.6. Organization capital as a substitute for measured inputs. Panel (a): elasticities of purchased inputs and of sales per worker with respect to B, as functions of the substitution share s, from (E.14). Panel (b): the substitution share over the firm life cycle, normalized so that $s = \psi$ in mid-life.

post-adoption dynamics governed by a single exact recursion. Two implications organize the empirical work. Adjustment is slow for mature firms not because of frictions but because their stocks are large relative to their flows: plausible depreciation rates deliver half-lives of five years or more, matching the sluggish responses we estimate. And the response is governed by the organization capital ratio: firms that are large relative to what they have codified respond more, respond faster, and gain most from adopting—which is why the heterogeneity test we take to the data conditions on the pre-adoption level of organization capital to size.