Who Hires for AI after ChatGPT? Evidence from Job Postings by U.S. Firms

By Zihan Sun*

ChatGPT made advanced AI tools easier to access, but it is unclear whether this translated into greater demand for AI-related workers. This paper studies how U.S. firms changed their demand for AI-related labor after the release of ChatGPT. Using Revelio Labs job postings from January 2021 to September 2025, I classify job descriptions into AI R&D, AI application, and non-AI roles, and aggregate expected hires to the firm-month level. This distinction separates hiring for internal AI capability building from hiring for application-oriented integration. I then compare post-ChatGPT changes across large and small firms in technology and non-technology sectors using panel fixed-effects regressions. The results show that AI labor demand did not expand uniformly. AI R&D hiring increased mainly among large technology firms, while AI application hiring also rose among large non-technology firms. Smaller firms show little evidence of increased AI hiring, and I find little evidence that AI-related hiring crowded out non-AI vacancies in the short run. I interpret these patterns as evidence that firms responded to easier access to AI along different labor-market margins: application-oriented hiring broadened among large firms, while internal AI capability building remained concentrated among firms with scale and technological capacity. These patterns help explain why broader access to AI tools need not translate into broad-based AI-specific hiring.

The release of ChatGPT in late 2022 made advanced AI capabilities much easier for firms to access. This change created an ambiguous prediction for labor demand. Easier access to AI may increase hiring if firms need workers to develop models, build AI systems, or integrate AI tools into products and workflows. At the same time, it may reduce the need for internal AI capability building if firms can rely on external tools, APIs, cloud services, or vendor-provided software. This paper asks whether firms increased their demand for AI-related labor after ChatGPT, which types of firms responded, and whether the response reflected AI development or application-oriented integration.

This distinction is important because job postings observe only one margin of AI adoption. A firm can use AI without hiring AI workers, especially if it relies

* Zihan Sun: Department of Economics, University of Texas at Austin (email: sunzh32@utexas.edu). This paper uses Revelio Labs job postings data and individual profiles data accessed through Wharton Research Data Services (WRDS). Portions of this research were conducted using computing resources provided by Texas A&M University.

on software subscriptions, APIs, cloud services, or vendor-provided tools. This is likely to matter for smaller firms, for which building an internal AI team may remain costly even after ChatGPT lowered access cost. A limited hiring response among smaller firms therefore should not be read as evidence that they did not use AI. It may instead indicate that AI adoption occurred through margins that are less visible in vacancy data. By contrast, when a firm posts vacancies for AI workers, adoption involves a more visible organizational investment in talent, data, infrastructure, and workflow redesign. I therefore interpret AI-related postings as evidence on whether firms build or integrate AI capabilities through internal labor, not as a complete measure of AI use.

Existing research has explored three related questions. The first question is whether AI raises productivity. Micro-level studies of specific tasks often find that generative AI can improve performance, reduce completion time, or help less experienced workers perform better (Noy and Zhang, 2023; Brynjolfsson, Li and Raymond, 2025). These studies show that AI can be productive in well-defined settings. At broader levels of aggregation, however, the evidence remains less settled. Existing estimates suggest that measured productivity gains appear positive but still modest, and the extent to which AI will generate large, broad-based productivity improvements remains uncertain (Acemoglu, 2025).

The second question is how AI affects labor. Much of this literature uses task-based or occupational measures of AI exposure and links them to employment, wages, or worker-level outcomes. These studies provide useful evidence on where AI may substitute for or complement labor (Frey and Osborne, 2017; Felten, Raj and Seamans, 2021; Eloundou et al., 2024). Some evidence suggests weaker subsequent labor demand in more exposed tasks or worker groups (Hosseini Maasoum and Lichtinger, 2025; Brynjolfsson, Chandar and Chen, 2025), but broad firm-level or industry-level labor replacement remains harder to detect (Acemoglu et al., 2022*a*; Hartley et al., 2024). Possible explanations are that AI changes the composition of work within jobs, increases demand for complementary workers, or creates new roles, even when some tasks are exposed to automation (Acemoglu and Restrepo, 2018; Hampole et al., 2025; Autor and Thompson, 2025; Manning and Aguirre, 2026).

The third question is which firms adopt or invest in AI. Existing evidence shows substantial heterogeneity across firms: AI adoption is more common among larger firms, younger firms, technology-intensive firms, and firms with stronger digital, organizational, or innovation-related complements (Acemoglu et al., 2022*b*, 2023; McElheran et al., 2024; Baslandze et al., 2026). This literature also suggests that adoption is not a uniform margin: firms differ not only in whether they use AI, but also in how AI is embedded in production. What remains less clear is how these firm-level differences in AI adoption translate into labor demand after a large decline in the cost of accessing AI.

This paper studies a narrower but economically important question: when access to AI becomes easier, which firms choose to build or integrate AI capabili-

ties through internal labor? This approach complements studies of AI exposure, productivity, and survey-based adoption. Exposure measures describe where AI could affect tasks, and surveys can measure whether firms report using AI. Vacancy data instead show whether firms are willing to devote workers to AI-related activities. This labor-market perspective allows me to separate two margins that are often combined in broader measures of AI labor: AI R&D roles, which are closer to building AI capabilities, and AI application roles, which are closer to integrating existing tools into products and workflows.

I use Revelio Labs job postings from January 2021 through September 2025 to measure firms' demand for AI-related labor. The sample consists of active online-posting firms that can be linked to firm characteristics, yielding a firm-month panel of 2,888 firms. This setting is well-suited for studying early firm responses to ChatGPT because job postings capture vacancy demand before changes appear in realized employment. Job descriptions also provide rich information about the tasks and responsibilities associated with each vacancy, allowing me to distinguish different forms of AI-related work within the same firm and even within the same role title.

I classify postings into AI R&D, AI application, and non-AI roles using a two-step procedure. I first use dictionaries to identify candidate AI-related postings, and then classify job descriptions based on role-specific evidence of AI work. This approach first differs from broader AI-skilled labor measures that often treat AI labor as a single category (Babina et al., 2024; Li, 2025). It also differs from widely used AI exposure measures. Exposure measures are useful for studying potential substitution or augmentation, but they capture potential changes to work tasks rather than actual AI use and do not imply inevitable displacement (Babina, 2026; Manning and Aguirre, 2026). They can also be sensitive to the rubric, model, or annotator used to construct exposure scores (Yin, Vu and Persico, 2026). My measure instead separates AI R&D from AI application roles and uses observed vacancy text to identify whether a firm is hiring for AI development or AI application work. It is also complementary to firm surveys, which can ask directly about AI use but may provide less information on the specific labor margin through which firms build or integrate AI internally. To reduce classification error, I validate the classification using manual checks and alternative large language model classifications.

The empirical analysis proceeds in three steps. To study heterogeneity, I classify firms along two predetermined dimensions: industry and pre-ChatGPT employment size. Combining these dimensions yields four groups: large technology, large non-technology, small technology, and small non-technology firms. I first describe raw pre- and post-ChatGPT changes in AI expected hires across these groups. I then estimate absolute within-group post changes using firm-month regressions with firm fixed effects and controls for seasonality. These regressions ask whether each firm group increased AI hiring after ChatGPT relative to its own pre-period level. Finally, I estimate the main relative specification with firm fixed effects

and monthly fixed effects. In this specification, small non-technology firms are the omitted group, and the coefficients measure whether other firm types expanded AI hiring more than small non-technology firms. The main outcomes are monthly $\log(1 + \text{expected hires})$ for AI R&D and AI application roles. I also examine non-AI and total expected hires to study short-run crowd-out at the vacancy margin.

The results show that post-ChatGPT AI labor demand increased unevenly across AI margins and across firms. The descriptive patterns and absolute within-group regressions show that AI R&D hiring increased significantly only among large technology firms, while AI application hiring increased among both large technology and large non-technology firms. The relative specifications sharpen this pattern. Compared with small non-technology firms, large technology firms increased both AI R&D and AI application hiring, while large non-technology firms increased AI application hiring but not AI R&D hiring. Small technology firms did not show a statistically significant increase in either AI margin. These results suggest that ChatGPT broadened the application margin of AI labor demand, but AI capability building remained concentrated among large technology firms. More broadly, this pattern points to one reason aggregate employment effects may be difficult to detect: firms may respond through different adoption margins, including margins that generate little new AI-specific hiring. The main AI-hiring findings are robust to rerunning the classification with an alternative LLM and to alternative definitions of technology firms, firm size, and the analysis sample.

I find little evidence of short-run crowd-out from AI-related hiring to non-AI hiring at the vacancy margin. Large technology firms increased AI-related hiring without a significant decline in non-AI or total expected hires, while large non-technology firms increased AI application hiring alongside broader vacancy expansion. These patterns are more consistent with continued demand for complementary workers and organizational capacity than with an immediate reallocation away from non-AI vacancies.

This paper makes three contributions. First, it provides a job-description-based measure of AI labor demand that distinguishes AI R&D from AI application roles. This distinction is useful because building AI capabilities and applying existing AI tools require different types of workers and are likely to respond differently to a decline in access costs.

Second, the paper clarifies the distinction between easier access to AI tools and internal AI capability building. ChatGPT may have lowered the cost of accessing powerful AI capabilities, but this does not imply that all firms have the incentives or capacity to build AI teams. The evidence shows that AI R&D hiring remained concentrated among large technology firms, while AI application hiring diffused more broadly but mainly among large firms. In this sense, the paper shows that democratizing access to AI tools is not the same as democratizing the organizational capacity to build and reorganize around AI.

Third, the paper provides a labor-demand interpretation of heterogeneous AI adoption. If firms respond through different adoption margins, then expanding AI use need not generate broad-based AI-related hiring. Weak AI-hiring responses among some firms should therefore be interpreted cautiously: they may reflect limited AI use, but they may also reflect adoption through software, APIs, vendors, or existing workers rather than new AI vacancies. This helps explain why aggregate employment effects may be difficult to detect even when AI use is expanding. Consistent with this interpretation, I find little evidence that early AI-related hiring came at the expense of broad non-AI vacancy demand.

These findings also have implications for interpreting aggregate labor-market effects. If AI use, AI application, and AI capability building occur through different firm-level margins, then the aggregate effects of AI will depend not only on what the technology can do, but also on which firms build, integrate, or purchase AI capabilities. The firm-level heterogeneity documented here suggests that muted aggregate employment responses need not imply limited AI use. They may instead reflect the fact that some forms of adoption require internal AI labor, while others are less visible in hiring data.

I. Data, Measurement, and Classification

This section describes how I construct the firm-month panel used to measure AI-related labor demand. The analysis combines online job postings, job descriptions, and firm characteristics. The goal is to convert job-level text into firm-month measures of AI R&D hiring, AI application hiring, non-AI hiring, and total expected hiring.

The main measurement challenge is that AI-related work can appear in different forms. Some jobs are closer to developing AI capabilities, while others are closer to integrating existing AI tools into products and workflows. These two types of AI labor demand have different economic meanings and are not always distinguishable from job titles alone. I therefore use a two-step classification procedure: dictionary-based candidate generation followed by job-description-level classification.

A. Job Postings Data and Sample Construction

The main data source is Revelio Labs job postings data for U.S. firms. For each job posting, the data contain a job identifier, the posting firm identifier, the posting date, the job title, the job description, and the expected number of hires. I use the firm identifier, `rcid`, to link postings to firm-level characteristics, including industry and pre-ChatGPT employment size.

The sample should not be interpreted as the universe of all U.S. firms. It consists of firms with observable online job postings, valid firm identifiers, and sufficient firm-level information. Relative to the broader population of firms, these firms are likely to be larger, more active in formal labor markets, and more likely to use

digital recruiting channels. The analysis therefore speaks to AI-related vacancy demand among active online-posting firms rather than all businesses.

The baseline sample covers January 2021 through September 2025. I start the sample in 2021 to avoid the sharp labor market disruptions during the first year of the COVID-19 pandemic. The sample window includes both the pre-ChatGPT period and the period after ChatGPT was publicly released in November 2022. In the baseline empirical specification, I define the post-ChatGPT period as beginning in December 2022.

The unit of observation in the main panel is a firm-month. I restrict the sample to firms that post at least ten jobs in at least one month during the sample period. This restriction does not require a firm to post ten jobs every month. It removes firms with too little observed posting activity to construct meaningful monthly outcomes. The resulting baseline sample contains 2,888 firms.

Job postings are useful for this research question because they measure vacancy demand and contain job-level text. They capture firms' forward-looking labor demand before changes appear in realized employment, which helps study early responses to the release of ChatGPT. Job descriptions also provide information about role-specific responsibilities and required skills, making it possible to distinguish jobs that develop AI systems from jobs that apply AI technologies or do not involve AI work. This job-description-based classification is less speculative than estimating occupational AI exposure: rather than asking whether AI could automate or accelerate a task, I classify whether the observed job description contains role-specific evidence of AI development, AI application, or neither.

A corresponding limitation is that job postings do not measure realized employment outcomes. Throughout the paper, outcomes based on expected hires should be interpreted as measures of vacancy demand or intended hiring.

B. Firm Type Classification

The empirical analysis compares how different types of firms adjust AI-related hiring after ChatGPT. I classify firms along two dimensions: industry and size. The industry dimension separates technology firms from non-technology firms. The size dimension separates large firms from smaller firms using the pre-ChatGPT employment size. Combining these two dimensions produces four firm groups: large technology firms, small technology firms, large non-technology firms, and small non-technology firms.

I define technology firms using NAICS codes. This classification is based on relatively stable pre-existing firm characteristics rather than post-ChatGPT outcomes. The technology industries include hardware manufacturing, software publishing, cloud infrastructure, data platforms, and IT services, corresponding to NAICS 334, 511, 518, 519, and 5415. Firms outside these industries are classified as non-technology firms.

Firm size is measured using pre-sample employment stock from Revelio individual profiles. I use February 2020 employment to avoid measuring firm size

after the pandemic shock or after the release of ChatGPT. Firms above the 75th percentile of the employment distribution are classified as large firms, and the remaining firms are classified as small firms. In the baseline sample, the 75th percentile cutoff corresponds to approximately 1,680 employees.

This four-group classification is motivated by the idea that firms may differ in both their ability and their need to build AI capabilities. Technology firms are more likely to have prior AI-related technical capacity and business exposure to AI technologies. Large firms are more likely to have the organizational resources, data infrastructure, and complementary assets needed to absorb the fixed costs of AI adoption. ChatGPT may lower the cost of accessing frontier AI capabilities, but firms may still respond in different ways. Large technology firms may expand AI development. Non-technology firms may focus more on AI application roles. Smaller firms may adopt AI through external software or subscriptions without building large internal AI teams.

The distribution of firms reflects the selected sample of active online-posting firms. Small non-technology firms are the largest group, with 1,826 firms, or about 63% of the sample. Large non-technology firms account for 710 firms. Technology firms are less common: the sample contains 239 small technology firms and 113 large technology firms. This distribution is useful for interpretation because the baseline group in many specifications is small non-technology firms, while large technology firms represent the group with the strongest prior AI-related capabilities.

C. Identifying AI-Related Job Postings

The central measurement task is to classify job postings into AI-related and non-AI categories. I use a two-step procedure. The first step uses dictionaries to generate a high-recall set of candidate AI-related postings. The second step uses a large language model classifier, implemented with ChatGPT 5.4 nano, to assign each candidate job description to one of four labels: `AI_RND`, `AI_APP`, `NON_AI`, or `UNCLEAR`. Appendix A reports the dictionary terms, classification prompt, manual audit procedure, and robustness checks using an alternative LLM classifier.

The dictionary step is used for candidate generation. It removes clearly unrelated postings from the classification task and reduces the number of job descriptions that require detailed review. This step is useful because most job postings do not contain AI-related work. At the same time, a purely dictionary-based approach would be too rigid for the final measure. It could classify a job as AI-related because it contains a small number of AI-related words, even if AI is peripheral to the job. It could also confuse ordinary software engineering, data analysis, cloud engineering, or business intelligence work with AI work. For this reason, I use dictionaries only to define the candidate set. Final labels are assigned using the job description.

The second-stage classifier reads the job description and assigns one of four labels:

`AI_RND`. This label refers to roles whose core responsibilities involve AI development or the AI model lifecycle. These include jobs that develop, train, fine-tune, evaluate, deploy, maintain, scale, monitor, or serve AI, machine learning, LLM, NLP, computer vision, reinforcement learning, MLOps, inference, or model infrastructure systems.

`AI_APP`. This label refers to roles whose core responsibilities involve applying, integrating, configuring, implementing, or delivering AI technologies in products, workflows, customer solutions, analytics systems, or decision systems. These roles do not own the underlying model development or AI infrastructure.

`NON_AI`. This label refers to roles without substantive AI, machine learning, or LLM work. Ordinary software engineering, data analysis, business intelligence, cloud engineering, generic automation, finance, operations, sales, teaching, maintenance, and healthcare roles are classified as non-AI unless the job description contains explicit job-specific AI responsibilities.

`UNCLEAR`. This label refers to postings with missing, empty, too short, or too vague descriptions. I use this label only when the description does not contain enough information to identify the actual duties of the job.

The classification follows three additional rules. First, the classifier focuses on job-specific responsibilities and actual duties. Since job descriptions often contain company overviews, product descriptions, or mission statements, a posting is not classified as AI-related only because the company describes itself as using AI. Second, the distinction between `AI_RND` and `AI_APP` depends on whether the job owns AI technology creation or applies existing AI capabilities. Third, the classification is conservative. If explicit job-specific AI evidence is absent, the posting is not classified as AI-related, even if the title contains AI-related terms.

Postings that are not selected by the dictionary stage are treated as non-AI in the main construction. Among candidate postings, only those classified as `AI_RND` or `AI_APP` enter the main AI hiring measures. Since most `UNCLEAR` labels are caused by insufficient information in the job description rather than substantively ambiguous AI postings, those classified as `UNCLEAR` are not counted as AI hiring either.

D. Outcome Variables

I aggregate classified postings to the firm-month level. The main outcomes measure AI R&D hiring, AI application hiring, total AI hiring, non-AI hiring, and total expected hiring.

Let H_{it} denote total expected hires by firm i in month t. Let H_{it}^{RND} denote expected hires from postings classified as `AI_RND`, and let H_{it}^{APP} denote expected hires from postings classified as `AI_APP`. Total AI hiring is defined as

$$H_{it}^{AI} = H_{it}^{RND} + H_{it}^{APP}.$$

Non-AI hiring is defined as residual hiring demand not classified as AI R&D or

AI application:

$$H_{it}^{NONAI} = H_{it} - H_{it}^{AI}.$$

The main AI measures are therefore conservative. A posting contributes to AI hiring only if it contains explicit job-specific evidence of AI R&D or AI application work.

The main hiring-volume outcomes are:

$$\log(1+H_{it}^{RND}), \quad \log(1+H_{it}^{APP}), \quad \log(1+H_{it}^{AI}), \quad \log(1+H_{it}^{NONAI}), \quad \log(1+H_{it}).$$

These log expected-hire measures capture changes in the level of vacancy demand and are the main outcomes in the empirical analysis. As supplementary measures of hiring composition, I also construct AI hiring shares, defined as AI expected hires divided by total expected hires at the firm-month level. Because changes in these shares can reflect either changes in AI hiring or changes in non-AI hiring, the main analysis emphasizes level outcomes, and the appendix reports share-based results.

II. Conceptual Framework and Empirical Implications

In this section, I use the distinction between AI capability building and AI application-oriented integration to guide the empirical analysis. The key idea is that lower access costs after ChatGPT need not affect all firms or all types of AI work in the same way. Firms differ in whether they build AI internally, adapt existing tools, or rely on external providers, and these choices depend on complementary assets such as scale, technical labor, data, and organizational capacity. This framework motivates estimating AI R&D and AI application hiring separately and comparing responses across firm types.

A. Adoption Margins and What Vacancy Data Observe

AI adoption can occur through different margins. Firms may build AI capabilities internally, integrate existing AI tools into firm-specific products and workflows, or rely on external providers. Job postings observe one part of this process: whether firms use labor to build or integrate AI capabilities.

The release of ChatGPT makes this labor margin especially relevant. By making advanced AI capabilities easier to access through a general-purpose interface, ChatGPT may have expanded firms' ability to experiment with AI without requiring every firm to build an internal AI team (Cai et al., 2025). Some responses require new AI-related workers, while others can occur through external tools or existing organizational routines. In vacancy data, only the hiring-based response is directly visible.

For the empirical analysis, this logic motivates estimating AI R&D and AI application hiring separately. AI R&D is closer to building AI capabilities, while AI application is closer to integrating existing AI tools into firm-specific products

and workflows. Both are observed in job postings, but they correspond to different organizational choices: hiring workers to build AI capabilities versus hiring workers to apply existing tools.

B. Firm Heterogeneity in the Make-or-Buy Margin

The make-or-buy margin is likely to vary systematically across firms. Large firms may have stronger incentives and greater capacity to build AI capabilities internally or customize AI tools to their own products, data, and workflows. Smaller firms may instead find it less costly to access AI through externally provided tools. Recent evidence on AI spending and software use is consistent with this margin, showing that firms can access AI through both internal development and externally provided software or services (de Souza, 2026; Baslandze et al., 2026).

This margin helps interpret firm-size differences in AI labor demand. A weak AI-hiring response among smaller firms does not necessarily imply a weak response to AI. It is also consistent with AI use that leaves a smaller footprint in vacancy data, for example when firms rely more on external tools or existing workers than on building new AI teams. Conversely, a stronger AI-hiring response among larger firms is more likely to reflect AI adoption that requires internal labor, such as capability building or firm-specific customization.

Lower access costs also do not eliminate the importance of complementary assets. Building or customizing AI capabilities requires more than access to a model. Firms also need specialized workers, relevant data, and use cases that justify reorganizing workflows around the technology (Aghion et al., 2019). These complements are likely to be stronger in large firms and technology-intensive firms. They should matter especially for AI R&D, which requires deeper technical capacity and specialized labor. AI application hiring may diffuse more broadly because it is closer to the use and integration margin. Even so, firms still need the capacity to identify use cases and embed AI tools into existing workflows.

C. Empirical Implications

The framework yields three empirical implications.

First, post-ChatGPT AI hiring should increase among firms with stronger complementary assets, especially firms that combine scale with technological capacity. If lower access costs make AI easier to use but internal capability building still requires technical labor, data, managerial capacity, and workflow integration, then the hiring response should be strongest among firms that already have these complements.

Second, AI application hiring may diffuse more broadly than AI R&D hiring, but this diffusion need not extend equally to smaller firms. AI R&D is closer to internal technology development and should depend more heavily on technical capacity. AI application is closer to the use and integration margin, making

it more likely than AI R&D to appear outside the technology sector. But it still requires firms to identify use cases, coordinate implementation, and embed AI tools into production. This makes it possible for the application margin to broaden beyond large technology firms without becoming broad-based across all firm groups.

Third, AI-related hiring need not immediately come at the expense of non-AI hiring. If building or integrating AI requires complementary workers, supports product development, or expands AI-enabled business functions, then firms increasing AI-related hiring may also continue hiring non-AI workers. I therefore examine non-AI and total hiring outcomes to ask whether AI-related hiring is associated with short-run crowd-out at the vacancy margin.

These implications guide the empirical analysis in the next section. I estimate whether post-ChatGPT changes in AI-related expected hires differ across the four firm groups defined above, focusing on level outcomes as the main measures of AI labor demand. I then examine non-AI and total expected hires to assess whether AI-related hiring is associated with short-run crowd-out at the vacancy margin.

III. Empirical Analysis

A. Descriptive Patterns and Within-Group Changes

Table 1—Pre- and Post-ChatGPT AI Expected Hires by Firm Group

Outcome	Firm group	Pre Mean	Post Mean	Change Post–Pre	Percent Change (%)
AI R&D	Large technology	0.474	0.543	0.070	14.7
AI R&D	Large non-technology	0.239	0.235	-0.003	-1.4
AI R&D	Small technology	0.079	0.066	-0.012	-15.6
AI R&D	Small non-technology	0.056	0.050	-0.006	-10.5
AI application	Large technology	0.281	0.377	0.096	34.2
AI application	Large non-technology	0.144	0.169	0.025	17.2
AI application	Small technology	0.036	0.044	0.008	22.5
AI application	Small non-technology	0.030	0.031	0.002	5.5

Note: The table reports firm-level averages of monthly log(1 + expected hires) within the pre- and post-ChatGPT periods. Each firm is weighted equally within a firm group. The four firm groups are defined in Section I.B. The pre-period ends before the release of ChatGPT, and the post-period begins after the release of ChatGPT.

Table 1 reports pre- and post-ChatGPT averages of AI expected hires by firm group. The table focuses on count-based outcomes, measured as monthly log(1 + expected hires). Two facts stand out. First, AI hiring was already highly uneven before ChatGPT: large technology firms had the highest pre-period levels of both AI R&D and AI application hiring, followed by large non-technology

firms. Second, the post-ChatGPT changes differ sharply across the two AI margins. AI R&D hiring increases only among large technology firms, while AI application hiring rises more broadly, especially among large technology and large non-technology firms.

These raw averages do not control for persistent differences across firms or seasonal hiring patterns. I therefore estimate a pooling regression that compares each firm's post-ChatGPT AI hiring demand with its own pre-period level, net of seasonality:

$$Y_{it} = \alpha_i + \eta_{m(t)} + \sum_{g \in G} \beta_g \left(Post_t \times FirmGroup_i^g \right) + \varepsilon_{it}, \tag{1}$$

$$G = \{LargeTech, LargeNonTech, SmallTech, SmallNonTech\}.$$

Here, i indexes firms and t indexes months. Firm fixed effects, α_i, absorb time-invariant differences across firms, including persistent differences in baseline hiring intensity, business models, and pre-existing technology orientation. Month-of-year fixed effects, $\eta_{m(t)}$, absorb recurring seasonal hiring patterns. $Post_t$ is an indicator for the post-ChatGPT period. The coefficients β_g estimate absolute post-ChatGPT changes within each firm group relative to that group's own pre-period level.

Table 2—Post-ChatGPT Changes in AI Expected Hires by Firm Group

	Log expected hires	
Firm group	AI R&D	AI application
Large technology	0.071**	0.097***
	(0.032)	(0.031)
Large non-technology	-0.003	0.025***
	(0.009)	(0.008)
Small technology	-0.010	0.009
	(0.010)	(0.007)
Small non-technology	-0.006*	0.002
	(0.003)	(0.002)
Firm FE	Yes	Yes
Month-of-year FE	Yes	Yes
Observations	161,342	161,342

Note: The table reports estimates from firm-month regressions of AI hiring outcomes on interactions between a post-ChatGPT indicator and the four firm-group indicators. The specification includes firm fixed effects and month-of-year fixed effects. The post period begins in December 2022. The four firm groups are defined in Section I.B. The dependent variables are monthly log(1+ expected hires) for AI R&D roles and AI application roles. The coefficients estimate absolute within-group post-ChatGPT changes. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

The regression results reinforce the descriptive patterns. AI R&D expected hires

increase significantly only among large technology firms. In contrast, AI application expected hires increase significantly among both large technology and large non-technology firms. Small firms do not show statistically significant increases in either log outcome. Appendix Table B1 reports the corresponding share-based estimates as supplementary evidence on hiring composition. The next subsection turns to relative differences across firm groups.

B. *Post-ChatGPT AI Hiring by Firm Type*

I estimate relative post-ChatGPT changes in AI hiring across the four firm groups. The baseline specification is

(2)
$$Y_{it} = \alpha_i + \lambda_t + \beta_1 Post_t \times LargeTech_i + \beta_2 Post_t \times LargeNonTech_i + \beta_3 Post_t \times SmallTech_i + \varepsilon_{it},$$

where the omitted group is small non-technology firms. The main outcomes are count-based measures of AI labor demand, $\log(1 + H^{RND}it)$ and $\log(1 + H^{APP}it)$. As before, firm fixed effects absorb time-invariant differences across firms. Unlike the within-group specification above, this specification includes monthly fixed effects, λ_t, which absorb aggregate labor-market conditions and other shocks common to all firms in a given month. The coefficients therefore measure differential post-ChatGPT changes in AI-related expected hires relative to small non-technology firms. Standard errors are clustered at the firm level.

Table 3—Post-ChatGPT AI Expected Hires by Firm Group

	AI R&D expected hires	AI application expected hires
Post × Large technology	0.0759**	0.0946***
	(0.0324)	(0.0306)
Post × Large non-technology	0.0022	0.0228***
	(0.0099)	(0.0082)
Post × Small technology	-0.0049	0.0068
	(0.0108)	(0.0070)
Firm fixed effects	Yes	Yes
Month fixed effects	Yes	Yes
Observations	161,342	161,342
R^2	0.710	0.630

Note: The dependent variables are monthly log(1 + expected hires) for AI R&D roles and AI application roles. The omitted group is small non-technology firms. Firm groups are defined in Section I.B. All specifications include firm fixed effects and month fixed effects. Standard errors are clustered at the firm level and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 3 reports the main estimates. Column 1 shows that the post-ChatGPT increase in AI R&D hiring is concentrated among large technology firms. Relative to small non-technology firms, large technology firms increase AI R&D expected

hires by 0.0759 log points after ChatGPT. The estimates for large non-technology firms and small technology firms are close to zero and statistically insignificant.

Pairwise contrasts, reported in Appendix Table B2, reinforce this interpretation. The large-technology response in AI R&D is significantly larger than the large-non-technology response and also significantly larger than the small-technology response. Thus, the AI R&D response is not simply a large-firm effect or a technology-sector effect in isolation. It is concentrated among firms that combine organizational scale with technological capability.

Column 2 shows a broader, but still uneven, response for AI application hiring. Large technology firms again exhibit the largest increase, with a coefficient of 0.0946 log points. Large non-technology firms also increase AI application expected hires by 0.0228 log points, while small technology firms do not show a statistically significant increase. Pairwise contrasts show that the large-technology response is significantly larger than both the large-non-technology and small-technology responses.

These results support the framework's main implication: post-ChatGPT AI labor demand expanded through different margins across firms. AI capability building, measured by AI R&D hiring, remained concentrated among firms with both scale and technological capacity. AI application hiring diffused more broadly, extending to large non-technology firms, but did not expand significantly among small firms.

C. Non-AI Hiring and Crowd-Out

Given the increase in AI-related hiring among large technology firms and large non-technology firms, I next examine whether these increases were accompanied by declines in non-AI hiring. This exercise provides evidence on short-run crowd-out at the vacancy margin.

I estimate absolute within-group changes using the same pooled within-group specification as in equation (1), replacing the dependent variable with non-AI expected hires and total expected hires. This specification asks whether each firm group reduced non-AI or total vacancy demand after ChatGPT relative to its own pre-period level.

Table 4 reports the results. For large technology firms, the estimates show no statistically significant decline in either non-AI or total expected hires, despite the significant increase in AI-related hiring documented above. For large non-technology firms, both non-AI and total expected hires increase significantly, suggesting that the rise in AI application hiring occurred alongside broader vacancy expansion rather than through a reduction in non-AI roles. Small technology firms show a significant decline in both non-AI and total expected hires, but this group does not exhibit a significant increase in AI hiring. This makes it difficult to interpret the decline as AI-driven crowd-out. Overall, I find little evidence that early AI-related hiring came at the expense of non-AI vacancy demand. Among the firm groups that increased AI hiring, broader vacancy demand

Table 4—Post-ChatGPT Changes in Non-AI and Total Expected Hires by Firm Group

	Log expected hires	
Firm group	Non-AI	Total
Large technology	-0.034	-0.032
	(0.079)	(0.079)
Large non-technology	0.362***	0.362***
	(0.033)	(0.033)
Small technology	-0.161***	-0.160***
	(0.059)	(0.059)
Small non-technology	-0.014	-0.014
	(0.020)	(0.020)
Firm FE	Yes	Yes
Month-of-year FE	Yes	Yes
Observations	161,342	161,342

Note: The table reports estimates from firm-month regressions of non-AI and total expected-hire outcomes on interactions between a post-ChatGPT indicator and the four firm-group indicators. The specification includes firm fixed effects and month-of-year fixed effects. The post period begins in December 2022. The four firm groups are defined in Section I.B. The dependent variables are monthly log(1+ expected hires) for non-AI roles and all roles. The coefficients estimate absolute within-group post-ChatGPT changes relative to each group's own pre-period level. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

was either stable or expanding.

D. Robustness

I report robustness checks in Appendix Tables B3–B7. I first examine whether the results depend on the large language model used in the second-stage job-description classification. Appendix Tables B3–B6 repeat the main analyses using the alternative LLM classifier, Gemini 2.5 Flash Lite, holding fixed the candidate universe, label definitions, and aggregation procedure. The alternative classifier assigns more candidate postings to AI categories and fewer to UNCLEAR, but the pattern of firm-level heterogeneity remains qualitatively unchanged. The AI R&D response remains concentrated among large technology firms. The AI application response remains strongest among large technology firms, with weaker positive evidence among large non-technology firms. Small technology firms again show no robust increase in either AI margin. These results suggest that the main conclusions are not driven by the choice of LLM classifier.

Appendix Table B7 reports additional checks that vary the firm-type definition and sample construction. I define technology firms using a narrower 4-digit NAICS classification, redefine large firms using the median of pre-period employment, and restrict the sample to a balanced panel of firms observed in every month from January 2021 to September 2025. These checks preserve the four-group specification and focus on the two main level outcomes, $\log(1 + H^{RND}it)$ and $\log(1 + H^{APP}it)$.

The main conclusions are stable across these additional checks. The post-ChatGPT increase in AI R&D hiring remains concentrated among large technology firms. For AI application hiring, the large-technology response is also robust. The evidence for large non-technology firms remains positive but is more sensitive to the firm-type definition: it is significant in the balanced-panel sample, marginally significant under the 4-digit technology definition, and smaller under the median-size cutoff. Small technology firms do not show a significant increase in either AI R&D or AI application hiring in any specification.

IV. Conclusion and Future Work

This paper studies how firms changed their demand for AI-related labor after the release of ChatGPT. Using job postings from U.S. firms, I distinguish between AI R&D roles, which are closer to AI capability building, and AI application roles, which are closer to integrating existing AI tools into products and workflows. The main finding is that post-ChatGPT AI labor demand did not expand uniformly across firms. AI R&D hiring increased primarily among large technology firms, while AI application hiring also rose among large non-technology firms. Smaller firms show little evidence of increased AI hiring.

Taken together, these patterns point to a more specific conclusion: lower access costs did not eliminate the firm-level barriers to building AI capabilities internally. ChatGPT made advanced AI tools easier to access, but building and customizing AI inside the firm still appears to require scale, specialized talent, and the organizational capacity to redesign workflows around the technology. The broader response in AI application hiring suggests that existing AI tools can be integrated outside the technology sector, but mainly among firms large enough to adapt them to firm-specific products and workflows.

The results also help interpret the labor-market effects of AI. AI adoption need not appear as broad-based AI-specific hiring. Some firms may build AI capabilities internally, others may hire workers to integrate existing tools, and still others may adopt AI through software, APIs, or vendor-provided tools that are less visible in vacancy data. A related result is that I find little evidence that early AI-related hiring came at the expense of non-AI vacancy demand. Among the firm groups that increased AI hiring, broader vacancy demand was either stable or expanding. This suggests that early AI-related hiring was accompanied by continued demand for complementary workers and organizational capacity, rather than by a simple reallocation away from non-AI jobs.

Future work should combine vacancy data with more direct measures of AI use, including software adoption, AI service purchases, APIs, cloud tools, and vendor products. Such data would make it possible to distinguish firms that build or integrate AI capabilities internally from firms that adopt AI primarily through external providers. A second direction is to study what firms use AI for and how those use cases affect workers. AI used for software development, customer service, forecasting, search, or workflow automation may have different

implications for tasks, wages, career paths, and organizational structure. These differences are likely to matter for understanding when AI complements labor, when it substitutes for particular tasks, and why aggregate employment effects may remain difficult to detect even as AI use expands.

REFERENCES

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

Acemoglu, Daron, and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American economic review*, 108(6): 1488–1542.

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

Acemoglu, Daron, Gary Anderson, David Beede, Catherine Buffington, Eric Childress, Emin Dinlersoz, Lucia Foster, Nathan Goldschlag, John Haltiwanger, Zachary Kroff, et al. 2023. "Advanced technology adoption: selection or causal effects?" Vol. 113, 210–214, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Acemoglu, Daron, Gary Anderson, David N Beede, Catherine Buffington, Eric E Childress, Emin Murat Dinlersoz, Lucia S Foster, Nathan Goldschlag, John C Haltiwanger, Zachary Kroff, et al. 2022*b*. *Automation and the workforce: A firm-level view from the 2019 annual business survey.* Vol. 30659, National Bureau of Economic Research.

Aghion, Philippe, Antonin Bergeaud, Richard W Blundell, and Rachel Griffith. 2019. "The innovation premium to soft skills in low-skilled occupations." *Available at SSRN 3489777*.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association*, 23(4): 1203–1271.

Babina, Tania. 2026. "Understanding Firms' AI Efforts and Their Economic Impact."

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics*, 151: 103745.

Baslandze, Salomé, Zachary Edwards, John Graham, Ty McClure, Brent H Meyer, Michael Sparks, Sonya R Waddell, and Daniel Weitz. 2026. "Artificial intelligence, productivity, and the workforce: Evidence from corporate executives." National Bureau of Economic Research.

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine?: Six Facts about the Recent Employment Effects of Artificial Intelligence."

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, 140(2): 889–942.

Cai, Junhui Jeff, Xian Gu, Liugang Sheng, Mengjia Xia, Linda Zhao, and Wu Zhu. 2025. "AI as Co-founder: GenAI for Entrepreneurship." *arXiv preprint arXiv:2512.06506*.

de Souza, Gustavo. 2026. "AI in the Office and the Factory: Evidence from Administrative Software Registry Data."

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. "GPTs are GPTs: Labor market impact potential of LLMs." *Science*, 384(6702): 1306–1308.

Felten, Edward, Manav Raj, and Robert Seamans. 2021. "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses." *Strategic management journal*, 42(12): 2195–2217.

Frey, Carl Benedikt, and Michael A Osborne. 2017. "The future of employment: How susceptible are jobs to computerisation?" *Technological forecasting and social change*, 114: 254–280.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence DW Schmidt, and Bryan Seegmiller. 2025. *Artificial intelligence and the labor market.* National Bureau of Economic Research.

Hartley, Jonathan, Filip Jolevski, Vitor Melo, and Brendan Moore. 2024. "The labor market effects of generative artificial intelligence." *Available at SSRN*.

Hosseini Maasoum, Seyed Mahdi, and Guy Lichtinger. 2025. "Generative AI as seniority-biased technological change: Evidence from US résumé and job posting data." *Available at SSRN 5425555*.

Li, Boyuan. 2025. "Ai washing." Working paper. Available from: https://bpb-use2. wpmucdn. com/sites ....

Manning, Sam J, and Tomás Aguirre. 2026. "How Adaptable Are American Workers to AI-Induced Job Displacement?" National Bureau of Economic Research.

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia Foster, and Nikolas Zolas. 2024. "AI adoption in America: Who, what, and where." *Journal of Economics & Management Strategy*, 33(2): 375–415.

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

Yin, Michelle, Hoa Vu, and Claudia Persico. 2026. "How (un) stable are LLM occupational exposure scores? evidence from multi-model replication." National Bureau of Economic Research.

Appendix A. Classification and Validation Details

A1. Dictionary-Based Candidate Generation

The dictionary stage is used only for candidate generation. It identifies postings that are potentially AI-related and sends them to the job-description-level classifier. The dictionary stage does not determine the final AI R&D or AI application labels. Final labels are assigned using the full job description, as described in Appendix A.2.

Candidate postings enter the classification universe through three inclusion sources: core AI title matches, strict job-description matches, and strong AI-adjacent title matches. I then apply a conservative hard-exclusion screen to remove obvious non-AI role titles before classification.

Step	Matching unit	Purpose
Core title dictionary	Standardized role title	Captures postings whose titles directly indicate AI, machine learning, LLM, NLP, computer vision, or related AI roles.
Strict JD dictionary	Job description	Adds postings whose titles are not explicit AI titles but whose job descriptions contain strict AI-related evidence. This step uses phrase-specific context rules to keep JD hits with stronger links to AI, model, or application work.
Strong adjacent title dictionary	Standardized role title	Captures adjacent technical titles that may involve AI work, such as applied scientist, algorithm, recommender, search relevance, robotics, perception, RPA, or MLOps roles.
Hard exclusion dictionary	Standardized role title	Removes obvious non-AI occupations and generic engineering titles unless the title contains explicit AI-related context.

Core title phrases include artificial intelligence, machine learning, deep learning, computer vision, natural language processing, large language model(s), generative AI, GenAI, foundation model(s), reinforcement learning, AI engineer, AI scientist, machine learning engineer, deep learning engineer, computer vision engineer, NLP engineer, LLM engineer, AI researcher, machine learning scientist, and research scientist machine learning. Short title tokens include AI, NLP, and LLM. Standalone ML and applied scientist are not included in the core title dictionary.

The strict job-description dictionary contains core AI concepts, AI model-lifecycle terms, AI tools and libraries, and application-oriented AI terms. Core concepts include deep learning, neural networks, natural language processing, reinforcement learning, large language models, foundation models, generative AI, diffusion models, transformer models, model serving, inference, AI models, AI systems, and AI algorithms. Model-lifecycle and tool terms include machine learning model training, model evaluation, model deployment, fine-tuning, MLOps, PyTorch, TensorFlow, Hugging Face, scikit-learn, OpenCV, and CUDA. Application-oriented terms include LLM APIs, OpenAI API, Anthropic API, retrieval-augmented generation, embedding search, vector databases, LangChain, LlamaIndex, AI agents, AI-powered products, AI implementation, AI integration, AI workflows, AI solutions, AI platforms, recommendation systems, search relevance, and fraud detection models.

For potentially ambiguous phrases, the strict JD dictionary requires local job-duty context. For example, machine learning must appear near model-building, training, deployment, evaluation, pipeline, inference, or similar work terms; LLM is retained only when it refers to language-model work rather than legal degrees; fine-tuning is retained only in model-related contexts; and AI platform or AI solution phrases are retained only when connected to implementation, integration, product, or workflow responsibilities.

The strong adjacent title dictionary includes applied scientist; algorithm engineer, algorithms engineer, and algorithm scientist; ranking engineer or scientist; search ranking and search relevance; relevance engineer or scientist; recommendation engineer, recommendations engineer, recommendation scientist, recommender systems, and recommendation systems; information retrieval; robotics engineer or scientist; autonomous systems and autonomous vehicle; perception engineer or scientist; sensor fusion; motion planning; path planning; intelligent automation; RPA developer, RPA engineer, and RPA architect; and MLOps. Boundary tokens include RPA and MLOps. Broad data and analytics titles, generic research scientist titles, generic modeling or optimization titles, and model-risk titles are not included in this adjacent-title dictionary.

The hard-exclusion dictionary removes obvious false-positive role titles. Exact exclusions include generic engineering titles such as engineer, engineer I, mechanical engineer, civil engineer, manufacturing engineer, process engineer, quality engineer, sales engineer, service engineer, support engineer, and related non-AI engineering titles. Additional pattern-based exclusions cover clearly non-AI oc-

cupations such as tutor, teacher, nurse, physician, therapist, counselor, dental, veterinary, pharmacy, driver, cashier, server, cook, chef, barista, bartender, and retail roles. These pattern-based exclusions are guarded by AI-context terms, so titles containing explicit AI, machine learning, NLP, LLM, GenAI, MLOps, robotics, autonomous, perception, RPA, or automation context are retained. Ambiguous implementation roles such as solutions engineer, solutions architect, technical consultant, and technical sales engineer are not hard-excluded at this stage; they remain in the candidate pool and are resolved by the job-description-level classifier.

A2. Job-Description-Level Classification Prompt

The second-stage classifier assigns each candidate posting to one of four labels: AI_RND, AI_APP, NON_AI, or UNCLEAR. I use ChatGPT 5.4 nano to classify the candidate job descriptions. The classifier uses the standardized role title and the job description, with the description truncated to 3,000 characters. The model is run in batches with temperature set to zero. The prompt instructs the model to focus on job-specific responsibilities rather than the role title alone or company boilerplate. The full static prompt is reproduced below.

```
You are a careful job-posting classifier.

You must return only one valid JSON object.
Do not return markdown, code fences, explanations, or any text outside the JSON objec

The JSON object must have exactly one top-level key: results.
results must be a list with exactly one result object for each input job.
Each result object must use exactly these keys:
job_id, label, confidence, evidence_flag, short_reason.
short_reason must be one concise sentence under 25 words.

Classify each job posting into exactly one of:
AI_RND, AI_APP, NON_AI, UNCLEAR.

You are classifying individual job descriptions for an economics research project on

Label definitions:

AI_RND:
The job's core responsibilities involve developing, training, fine-tuning, evaluating
deploying, or maintaining AI/ML/LLM/NLP/CV/RL models, algorithms, model infrastructur
MLOps, inference systems, or AI systems.

AI_APP:
```

```
The job's core responsibilities involve applying, integrating, configuring, or delive
AI/ML/LLM technologies into products, business processes, customer solutions, data pr
workflows, or decision systems. This includes building AI-powered applications, using
RAG, embeddings, vector databases, or AI-enabled product features.

NON_AI:
The job description does not contain substantive AI/ML/LLM work. Ordinary software en
data analysis, BI, SQL, dashboards, cloud engineering, generic automation, teaching,
retail, sales, finance, trading, operations, or maintenance should be NON_AI unless e
AI/ML/LLM responsibilities are present.

UNCLEAR:
The description is missing, too short, or too vague to classify.

Important rules:
- Classify the job description, not the role title alone. Focus on job-specific respo
- Label AI_RND or AI_APP only when AI/ML/LLM/NLP/CV/RL work is explicitly tied to thi
- Company boilerplate rule: If AI/ML/LLM appears only in the company overview, produc
- AI_RND vs AI_APP:
  AI_RND = the job owns AI/ML technology creation or model lifecycle work.
  AI_APP = the job applies, integrates, configures, or delivers existing AI/ML/LLM ca
- Classify as AI_RND when the job's own duties involve developing, training, fine-tun
- For technical ML engineer/scientist/developer roles, classify as AI_RND when the ro
- Classify as AI_APP when the job mainly applies, integrates, configures, deploys, in
- For customer-facing, consulting, product, project-management, solution-architect, c
- Deployment, monitoring, or productionization alone is not enough for AI_RND. It is
- Do not infer AI work merely from words like data, analytics, automation, digital, a
- Do not classify as AI_RND when the evidence is only AI-adjacent infrastructure, clo
- For non-AI technical jobs involving non-AI algorithms, models, firmware, optimizati
- If AI/ML appears only in preferred qualifications, classify as NON_AI unless it is
- Do not classify a job as AI_APP merely because the worker may use ChatGPT, Copilot,
- If the title is AI/ML-related but the description lacks job-specific AI/ML duties c
- Use UNCLEAR only when the description is missing, empty, too short, or too vague to
- Be conservative. If explicit job-specific AI/ML/LLM evidence is absent, choose NON_
- Use evidence_flag to indicate the main reason for the classification.

Allowed evidence_flag values:
explicit_ai_ml_model_development
ml_engineering_or_mLOps
llm_or_genai
nlp_cv_rl
ai_product_or_solution
ai_integration_or_workflow
```

```
ordinary_software_or_cloud
generic_data_or_analytics_only
generic_automation_only
non_ai_algorithm_or_modeling
company_boilerplate_only
preferred_only_ai_signal
non_ai_occupation
insufficient_information
```

For each input posting, the model returns one classification result with five fields: job_id, label, confidence, evidence_flag, and short_reason. The confidence score and evidence flag are used for validation and audit checks, while the final empirical measures only use the labels.

A3. Manual Audit and Validation

I used manual audit checks both to refine the prompt and to validate the final classifier. In an initial local pilot, I constructed a sample from two sources: boundary or ambiguous role titles and job descriptions containing AI-related dictionary terms. The boundary sample was designed to include roles where AI classification was likely to be difficult, while the description-positive sample was designed to test whether AI-related language in the text reflected actual job responsibilities.

The initial pilot used ChatGPT 5.4 mini as a prompt-development exercise. This pilot revealed an important source of false positives: some postings contained AI-related language in company descriptions, team overviews, product descriptions, or aspirational statements, even though the specific job did not involve AI work. I therefore revised the prompt to emphasize role-specific responsibilities, required qualifications, and actual duties. I also added an explicit company-boilerplate rule: if AI-related language appears only in the employer overview, product description, mission statement, or general company AI initiatives, the posting is classified as NON_AI rather than AI_RND or AI_APP.

The production classification used ChatGPT 5.4 nano. Before applying the classifier to the full candidate universe, I ran small-batch tests using the final model and prompt. After these checks, I applied the classifier to the candidate universe in batches. Each posting received a label, confidence score, evidence flag, and short reason. I then manually audited a final quality-check sample from the production output. In this final audit file, 5 of 1,152 reviewed rows required manual overrides. This corresponds to an override rate of 0.43 percent. The label distribution in this audit sample was 49.4 percent AI_APP, 11.9 percent AI_RND, 20.1 percent NON_AI, and 18.6 percent UNCLEAR.

A4. Alternative LLM Classification

As a robustness check, I repeat the second-stage job-description classification using Gemini 2.5 Flash Lite. This exercise changes only the LLM used for classi-

fication. The candidate universe, label definitions, static prompt, output schema, post-processing rules, and firm-month aggregation procedure are held fixed. The goal is to assess whether the empirical results depend on the specific LLM classifier used to convert candidate job descriptions into AI R&D, AI application, non-AI, and UNCLEAR labels.

The alternative classifier produces a somewhat more AI-inclusive label distribution than the baseline ChatGPT 5.4 nano classifier. Among the same 501,938 candidate postings, the total AI share rises from 66.9 percent to 74.0 percent. This increase is driven mainly by a higher AI R&D share, while the UNCLEAR share falls from 5.9 percent to 3.1 percent. Thus, the alternative classifier is more likely to resolve candidate postings into substantive AI categories rather than UNCLEAR or non-AI labels.

Appendix Tables B3–B6 report the corresponding firm-level results using the Gemini-based labels. Despite the shift in the job-level label distribution, the main empirical patterns are qualitatively unchanged. AI R&D hiring remains concentrated among large technology firms, and AI application hiring remains strongest among large firms. The robustness check therefore suggests that the paper's main conclusions are not driven by the choice of LLM classifier.

Appendix B. Additional Results and Robustness Tables

Table B1—Post-ChatGPT Changes in AI Hiring Shares by Firm Group

	Expected-hire share	
Firm group	AI R&D	AI application
Large technology	0.076**	0.076***
	(0.035)	(0.029)
Large non-technology	0.003	0.027***
	(0.012)	(0.005)
Small technology	0.012	0.059*
	(0.046)	(0.034)
Small non-technology	0.046**	0.022
	(0.022)	(0.014)
Firm FE	Yes	Yes
Month-of-year FE	Yes	Yes
Observations	161,342	161,342

Note: The table reports estimates from firm-month regressions of AI hiring-share outcomes on interactions between a post-ChatGPT indicator and the four firm-group indicators. The specification includes firm fixed effects and month-of-year fixed effects. The dependent variables are AI R&D expected hires and AI application expected hires divided by total expected hires at the firm-month level. Coefficients and standard errors are multiplied by 100 and can be interpreted as percentage-point changes. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table B2—Pairwise Contrasts Across Firm Groups

Outcome	Contrast	Estimate	Standard error
AI R&D	Large technology − Large non-technology	0.0736**	(0.0336)
AI R&D	Large technology − Small technology	0.0807**	(0.0339)
AI application	Large technology − Large non-technology	0.0719**	(0.0315)
AI application	Large technology − Small technology	0.0878***	(0.0312)

Note: The table reports pairwise contrasts from the four-group specification in Table 3. The dependent variables are monthly $\log(1 + \text{expected hires})$ for AI R&D roles and AI application roles. Each contrast compares the post-ChatGPT response of large technology firms with the response of another firm group. Standard errors are computed using the covariance matrix from the main regression and are reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B3—Alternative LLM Classifier: Pre- and Post-ChatGPT AI Expected Hires by Firm Group

Outcome	Firm group	Pre Mean	Post Mean	Change Post–Pre	Percent Change (%)
AI R&D	Large technology	0.501	0.586	0.085	16.9
AI R&D	Large non-technology	0.255	0.255	0.001	0.2
AI R&D	Small technology	0.084	0.072	-0.011	-13.5
AI R&D	Small non-technology	0.060	0.056	-0.004	-6.7
AI application	Large technology	0.303	0.378	0.076	25.0
AI application	Large non-technology	0.155	0.168	0.013	8.2
AI application	Small technology	0.043	0.045	0.002	4.8
AI application	Small non-technology	0.032	0.030	-0.003	-8.3

Note: The table repeats the descriptive pre- and post-ChatGPT averages using the alternative Gemini 2.5 Flash Lite classifier. The table reports firm-level averages of monthly $\log(1 + \text{expected hires})$ within the pre- and post-ChatGPT periods. Each firm is weighted equally within a firm group. The four firm groups are defined in Section I.B. The pre-period ends before the release of ChatGPT, and the post-period begins after the release of ChatGPT.

Table B4—Alternative LLM Classifier: Post-ChatGPT Changes in AI Expected Hires by Firm Group

	Log expected hires		Expected-hire share	
Firm group	AI R&D	AI application	AI R&D	AI application
Large technology	0.086**	0.077**	0.091**	0.087***
	(0.035)	(0.031)	(0.037)	(0.032)
Large non-technology	0.001	0.013*	0.008	0.021***
	(0.010)	(0.008)	(0.013)	(0.006)
Small technology	-0.009	0.003	0.030	0.032
	(0.010)	(0.008)	(0.045)	(0.040)
Small non-technology	-0.004	-0.002	0.062***	-0.008
	(0.003)	(0.003)	(0.021)	(0.019)
Firm FE	Yes	Yes	Yes	Yes
Month-of-year FE	Yes	Yes	Yes	Yes
Observations	161,342	161,342	161,342	161,342

Note: The table repeats the absolute within-group post-ChatGPT regressions using the alternative Gemini 2.5 Flash Lite classifier. The table reports estimates from firm-month regressions of AI hiring outcomes on interactions between a post-ChatGPT indicator and the four firm-group indicators. The specification includes firm fixed effects and month-of-year fixed effects. The post period begins in December 2022. The first two columns use monthly log(1+ expected hires) for AI R&D roles and AI application roles as dependent variables. The last two columns use AI R&D and AI application expected hires divided by total expected hires at the firm-month level; coefficients and standard errors in these share columns are multiplied by 100 and can be interpreted as percentage-point changes. The coefficients estimate absolute within-group post-ChatGPT changes. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

TABLE B5—ALTERNATIVE LLM CLASSIFIER: POST-CHATGPT AI EXPECTED HIRES BY FIRM GROUP

	Log expected hires	
Firm group	AI R&D	AI application
Post × Large technology	0.089**	0.078**
	(0.035)	(0.031)
Post × Large non-technology	0.004	0.015*
	(0.010)	(0.008)
Post × Small technology	-0.006	0.005
	(0.011)	(0.008)
Firm FE	Yes	Yes
Month FE	Yes	Yes
Observations	161,342	161,342
R^2	0.713	0.626

Note: The table repeats the main four-group specification using the alternative Gemini 2.5 Flash Lite classifier. The dependent variables are monthly log(1+ expected hires) for AI R&D roles and AI application roles. The omitted group is small non-technology firms. Firm groups are defined in Section I.B. All specifications include firm fixed effects and month fixed effects. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

TABLE B6—ALTERNATIVE LLM CLASSIFIER: POST-CHATGPT CHANGES IN NON-AI AND TOTAL EXPECTED HIRES BY FIRM GROUP

	Log expected hires	
Firm group	Non-AI	Total
Large technology	-0.034	-0.032
	(0.079)	(0.079)
Large non-technology	0.362***	0.362***
	(0.033)	(0.033)
Small technology	-0.161***	-0.160***
	(0.059)	(0.059)
Small non-technology	-0.014	-0.014
	(0.020)	(0.020)
Firm FE	Yes	Yes
Month-of-year FE	Yes	Yes
Observations	161,342	161,342

Note: The table repeats the non-AI and total expected-hire regressions using the alternative Gemini 2.5 Flash Lite classifier. The table reports estimates from firm-month regressions of non-AI and total expected-hire outcomes on interactions between a post-ChatGPT indicator and the four firm-group indicators. The specification includes firm fixed effects and month-of-year fixed effects. The post period begins in December 2022. The four firm groups are defined in Section I.B. The dependent variables are monthly log(1+ expected hires) for non-AI roles and all roles. The coefficients estimate absolute within-group post-ChatGPT changes relative to each group's own pre-period level. Standard errors, clustered at the firm level, are reported in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent levels.

Table B7—Post-ChatGPT AI Hiring under Alternative Firm Definitions and Samples

	AI R&D expected hires	AI application expected hires
Panel A. Narrow 4-digit technology definition		
Post × Large technology	0.0504*	0.0856***
	(0.0267)	(0.0229)
Post × Large non-technology	-0.0022	0.0150*
	(0.0097)	(0.0081)
Post × Small technology	-0.0128	-0.0064
	(0.0079)	(0.0064)
Observations	161,342	161,342
Panel B. Median size cutoff		
Post × Large technology	0.0443**	0.0685***
	(0.0221)	(0.0203)
Post × Large non-technology	-0.0047	0.0091
	(0.0069)	(0.0056)
Post × Small technology	-0.0095	-0.0043
	(0.0122)	(0.0073)
Observations	161,342	161,342
Panel C. Balanced panel		
Post × Large technology	0.0896***	0.1042***
	(0.0343)	(0.0328)
Post × Large non-technology	0.0062	0.0234***
	(0.0099)	(0.0084)
Post × Small technology	-0.0012	0.0043
	(0.0089)	(0.0066)
Observations	140,790	140,790
Firm fixed effects	Yes	Yes
Month fixed effects	Yes	Yes

Note: The robustness checks use three alternative samples or definitions. First, the 4-digit technology definition classifies firms as technology firms if their 4-digit NAICS code is 3341, 3342, 3344, 3345, 5132, 5182, 5192, or 5415. Second, the median-size specification defines large firms as firms with baseline employment proxy above the sample median, corresponding to a cutoff of 433. Third, the balanced-panel specification restricts the sample to firms observed in every month from January 2021 to September 2025. The dependent variables are monthly log(1 + expected hires) for AI R&D roles and AI application roles. Each panel reports estimates from the four-group specification. The omitted group is small non-technology firms under the corresponding firm-type definition. All specifications include firm fixed effects and month fixed effects. Standard errors are clustered at the firm level and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.