Prompted to Start: How Generative AI is Transforming Entrepreneurship*

Jan Bena† Bo Bian‡ Mariassunta Giannetti§

July 8, 2026

Abstract

We show that following the diffusion of GenAI, industries with higher task-level exposure to the new technology experience 20% more startup formation than less-exposed industries, and a growing share of these startups offer new AI products and services rather than merely automating operations. This entry extends beyond traditional innovation hubs into regions with limited venture capital and thin specialized labor markets. While individual entrants are smaller in scale, aggregate entry generates net employment and wage growth at the industry level and creates employment in the very occupations most exposed to GenAI. GenAI also changes who becomes an entrepreneur: founding teams are increasingly drawn from high-exposure occupations and emphasize their technical, AI-related skills rather than their capital-raising and deal-making expertise.

Keywords: *AI, Entrepreneurship, Startups, Founder, Agglomeration*

JEL classification: *J24, L26, M13, O33*

*We thank Ramana Nanda, David Robinson, and participants at the Adam Smith Conference, the Bank of Canada, the CEPR/CSEF Conference on Labour and Finance in Capri, the Collegio Carlo Alberto-Università di Torino, the NBER-SAIF Conference on AI and Financial Markets, the HEC Montreal Conference on Policy Challenges in a Changing Global Economy, the Norwegian School of Economics, the NOVA Entrepreneurship and Innovation Symposium, the Pontifical Catholic University of Chile, Skema International Workshop on Economic Growth and Macroeconomic Dynamics, and the Virtual Corporate Finance Seminar for their comments, and Yile Chen, Khaled Hanafi, and Andy Xin for excellent research assistance. We acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (Grant Number: 435-2023-0282), the Jan Wallander and Tom Hedelius Foundation (Grant Number: P25-0131), and the Stone Center on Wealth and Income Inequality. The usual disclaimer on errors applies.

†The University of British Columbia, Sauder School of Business. E-mail: jan.bena@sauder.ubc.ca.

‡The University of British Columbia, Sauder School of Business. E-mail: bo.bian@sauder.ubc.ca.

§Stockholm School of Economics, CEPR, & ECGI. E-mail: Mariassunta.Giannetti@hhs.se.

1 Introduction

Generative artificial intelligence (GenAI) is increasingly viewed as a general-purpose technology with the potential to reshape production and innovation through its broad applicability.[1] Accordingly, some observers emphasize the potential for substantial productivity gains and faster economic growth (Jones, 2024; Brynjolfsson et al., 2025). At the same time, critics argue that GenAI may automate cognitive tasks without generating commensurate productivity gains, raising concerns about labor displacement (Acemoglu et al., 2022; Felten et al., 2023). Existing evidence remains inconclusive, in part because the adoption and diffusion of new technologies are gradual.

We contribute to this debate by examining how GenAI affects entrepreneurial entry and the characteristics of new ventures. Entry decisions are inherently forward-looking. Therefore, the effects of a new general-purpose technology should surface earlier in new ventures than in macroeconomic outcomes or incumbent firms, which typically adopt more slowly. Because startups play a central role in job creation and reallocation (Haltiwanger et al., 2013; Decker et al., 2016), their organizational choices offer a leading indicator of broader economic impacts.

We consider two non-mutually exclusive channels through which GenAI can influence startup formation. First, by expanding the production possibility frontier, GenAI may enable new products and services. Startups whose core offerings rely on GenAI capabilities—such as AI-powered design tools, automated content generation, or conversational interfaces—represent entry aiming to exploit the new investment opportunities created by GenAI. Second, even when GenAI is peripheral to a firm's value proposition, it may reduce fixed entry costs by automating and augmenting non-core tasks such as prototyping, coding, marketing, and customer support, and enabling increases in output without a proportional increase in employment. This may reduce reliance on external finance and specialized labor and may weaken the geographic concentration of entrepreneurship in venture capital centers

[1] For classic discussions of general-purpose technologies and their implications for long-run growth, see Bresnahan and Trajtenberg (1995) and Jovanovic and Rousseau (2005).

(Chen and Ewens, 2025; Bennett and Robinson, 2024).

Countervailing forces may, however, limit entry. Developing frontier GenAI systems requires substantial investments in data, compute infrastructure, and specialized human capital, reinforcing incumbents' scale advantages.[2] Moreover, if GenAI primarily substitutes for labor without creating new market opportunities, it may reduce rather than expand entrepreneurial prospects. Our analysis of entry patterns, startup characteristics, and founder composition provides evidence on which forces dominate.

Our empirical strategy exploits variation in industries' exposure to GenAI based on the task content of occupations. We measure task-level exposure using the actual proportion of queries to GenAI systems for specific tasks, reflecting how tech-savvy early adopters use the technology. Following Autor et al. (2003), we aggregate task exposure to occupation and industry levels, yielding time-invariant measures that predict differential treatment intensity as GenAI diffuses. We then employ difference-in-differences specifications comparing entrepreneurial outcomes in high- versus low-exposure industries before and after GenAI's public release, beginning with Stable Diffusion in August 2022 and accelerating after ChatGPT's launch in November 2022.

We document four main findings. First, industries with high GenAI exposure experience a 20% increase in startup formation relative to low-exposure industries following ChatGPT's release. This effect is statistically significant starting in Q1 2023 and grows over time, indicating that GenAI induces substantial new entry rather than merely shifting the timing of marginal ventures. Using complementary LLM-assisted pipelines, we highlight two channels through which GenAI favors entrepreneurial entry: (i) "AI-enabled" startups automate and augment existing processes using GenAI, and (ii) "AI-native" ventures offer new products and services built on GenAI capabilities. The presence of AI-native startups suggests GenAI is creating new markets and products rather than simply reducing entry costs and may thus lead to exponential growth (Aghion and Howitt, 1992; Klette and Kortum, 2004).

[2] On rising market power and "superstar firm" dynamics, see Autor et al. (2020); De Loecker et al. (2020).

Second, GenAI-enabled entry is geographically dispersed beyond traditional innovation hubs. New startups in high-exposure industries increasingly locate in regions with limited venture capital access and thin specialized labor markets. This pattern is consistent with GenAI lowering entry costs for entrepreneurs and enabling leaner operations.

Third, while individual startups in high-exposure industries enter at a smaller scale, aggregate entry generates net employment growth. Higher entry rates more than offset smaller firm size, resulting in a 7% increase in employment and a 5% increase in average earnings in new firms in high-exposure GenAI industries.

Fourth, GenAI fundamentally alters the composition of labor demand within startups. Employment in new ventures concentrates in occupations whose tasks are highly exposed to GenAI, and founders in high-exposure industries are significantly more likely to have prior work experience in GenAI-exposed occupations, enabling them to combine domain expertise with the productivity gains the new technology offers. Thus, the skills demanded by startups differ from those demanded by established firms, where employment in GenAI-exposed occupations is declining (Acemoglu et al., 2022). Founders' self-described capabilities emphasize technical skills and de-emphasize capital-raising and deal-making skills, consistent with the finding that entry increasingly occurs outside venture capital hubs, and suggesting that GenAI reduces dependence on external finance. Founders come from companies that were expanding employment, have compensation comparable to founders of startups less exposed to GenAI, and do not appear to experience compensation cuts in their transition to entrepreneurship. These findings suggest that founders take advantage of the opportunities offered by the new technology and are not pushed into entrepreneurship by uncertain job prospects at their previous employers.

Overall, our results suggest that business dynamism can mediate the labor-demand contractions GenAI produces among incumbents and that entrepreneurial entry constitutes a previously neglected countervailing channel. Aggregate entry more than offsets startups' smaller scale, generating net employment and wage growth in new firms at the industry level. Furthermore, while established companies demand fewer workers in GenAI-exposed occupations (Acemoglu et al., 2022), startups hire precisely these workers to exploit the new opportunities the technology creates.

Our paper contributes to several strands of the literature. A growing body of work examines GenAI's labor-market implications. Theoretically, the implications for the demand of specific skills and wage inequality are ambiguous (Agrawal et al., 2026). Acemoglu et al. (2022) and Brynjolfsson et al. (2025) document increasing demand for AI skills and a decrease in non-AI hiring and GenAI-exposed occupations for established firms. These changes appear to be accompanied by productivity gains for less-exposed workers (Hampole et al., 2025). We document that founders emphasize their AI skills also in startups; however, in startups, AI skills appear to be associated with job creation in GenAI-exposed occupations, possibly contributing to the redesign of these jobs towards less GenAI-automable tasks (Brynjolfsson et al., 2018). In contemporaneous work, Gupta et al. (2026) study how GenAI reshapes employment, productivity, and financing within existing high-growth startups, documenting a 3% workforce contraction at exposed firms alongside accelerated scaling and fundraising. Our focus is different: we examine the extensive margin of entry and the reallocation of human capital toward new firm formation. Not only do we estimate a much larger 25% drop in the number of employees in GenAI-exposed startups entering following the GenAI diffusion, but we also show that GenAI changes founding teams' characteristics.

The literature on how GenAI shapes entrepreneurial entry is still in its early stages. Bao et al. (2025) focus on the tech sector, where GenAI may have pushed displaced individuals into entrepreneurship, while contemporaneous work by Marchesi and Tang (2026) document industry-level entry consistent with reduced experimentation costs, and Cai et al. (2025) report a post-ChatGPT rise in small-firm registrations in roughly 5 km^2 grid cells in China that had pre-2020 AI patents. We use a more general measure of exposure based on tasks actually performed through GenAI, document the mechanisms of technological diffusion that drive entry by measuring adoption directly, and assess the consequences of GenAI adoption for the internal organization of startups and the characteristics of founding teams.

Another strand of literature studies firm-level GenAI exposure in public companies. Eisfeldt et al. (2024) find that high-exposure firms earned excess returns following ChatGPT's release through technology-labor substitution. Babina et al. (2024) document positive effects of AI workforces on listed firms' growth and innovation, while field experiments show GenAI improves worker productivity within firms (Brynjolfsson et al., 2025). However, macro-level impacts on productivity and employment remain negligible (Bick et al., 2025; Humlum and Vestergaard, 2025). By focusing on entrepreneurial ventures, we capture early-stage economic effects and provide evidence that GenAI may eventually drive productivity growth through startup formation and business model innovation.

Finally, we contribute to research on how technological change affects industrial structure. Previous work shows that data and computing power favor concentration and incumbent advantages (Eeckhout and Veldkamp, 2022; Lu et al., 2024), though cloud computing reduces capital requirements for AI startups (Ewens et al., 2018). We show that GenAI tools—which can be adopted at low fixed cost—have opposite effects, promoting entry.

The paper proceeds as follows. Section 2 describes our data sources. Section 3 presents our measure of GenAI exposure. Section 4 describes our main results on entry patterns. Section 5 analyzes startup characteristics, including geographical location, size, and employment composition. Section Section 6 examines the characteristics of founding teams. Section 7.1 presents several robustness tests. Section 8 concludes.

2 Data and Measurement

We use a variety of data sources to measure entrepreneurial entry, firm dynamics, and startup characteristics.

2.1 Entrepreneurial Entry and Firm Dynamics

We begin with the U.S. Census Bureau's Business Formation Statistics (BFS). The BFS is constructed by the Census Bureau using administrative data from the Internal Revenue Service (IRS) on Employer Identification Number (EIN) applications. All employer businesses in the United States are required to have an EIN to file payroll taxes. New nonemployer businesses also file for an EIN if forming a partnership or an incorporated business. Even new sole proprietors often file for an EIN to facilitate their business activities (e.g., working with other businesses or opening a business bank account).

The Census Bureau publishes monthly the number of EINs applications submitted to the IRS at the two-digit NAICS level. Projected business formation statistics are computed by the U.S. Census Bureau using predictive models that link information from the EIN applications to subsequent business openings. Since data are published virtually in real time, they serve as leading indicators of actual firm creation across industries and are commonly used to provide early signals of nascent entrepreneurship (Dinlersoz et al., 2021).

This dataset allows us to track the dynamics of business formation in real time, but at the cost of a rather coarse industry aggregation. To capture industry dynamics with higher granularity, we use *Data Axle* (previously named InfoGroup), which tracks the universe of U.S. establishments and contains information on founding year, location, industry classification (NAICS), and employment. The database is constructed from multiple administrative and commercial sources, including business directories, state registration records, telephone and utility connections, and verified survey data, and is widely used in academic research (Acharya et al., 2023; Borisov et al., 2021; Denes et al., 2023). Data Axle's granularity enables analysis of new-firm creation, size distribution, and fine levels of industry aggregation (e.g., four-digit NAICS).

Table 1 reports summary statistics for the main variables. Panel A presents two-digit industry-year-month business formation statistics, while Panel B presents the four-digit industry-year dataset we construct from Data Axle for the period 2020 to 2024. At the four-digit level, while almost 80% of industry-years experience entry, the number of firms entering varies widely across industries. We also consider the number of entrants with sizes below/above the industry median at entry in a given year.

To assess the broader labor-market implications of changes in entrepreneurial entry, we complement the above data sources with the Census Quarterly Workforce Indicators (QWI). QWI provides a nationally consistent measure of employment and average monthly earnings of employees in firms of different ages at the industry (NAICS-4) and quarter level. In our empirical tests, we distinguish between very young firms (ages 0–1) and the rest of the companies to evaluate the employment effects of startups. Table D.1 provides summary statistics.

2.2 Startups and Founding Teams' Characteristics

To study labor demand in new firms, we construct an employer-employee-matched database linking Crunchbase startups to LinkedIn employee data.

We first obtain data from Crunchbase, a database on innovative companies, maintained by Crunchbase Inc., which is increasingly used by the venture capital industry to identify new investments. We consider five annual snapshots of startups founded in the United States between 2020 and 2024. From Crunchbase, we obtain each startup's proprietary industry classification and measure employment at the end of the founding year.

To obtain further information on the startups and their founding teams, we link firm information from Crunchbase to employee-level data from LinkedIn, via Revelio Labs, using the firm-specific LinkedIn URLs provided in Crunchbase. We identify employees who list the startup as their current employer on LinkedIn as of the end of the founding year. Each employee's job title, reported at the eighth-digit SOC level, is mapped to a six-digit occupation code using the U.S. Department of Labor's O*NET classification system, allowing us to estimate the number of employees in each occupation within the firm.Panel C of Table 1 summarizes the characteristics of all startups at the end of the founding year.

We further identify the founders of each startup using the founder records provided in Crunchbase. For each founder, we extract individual-level information, including professional background, prior work experience, and demographic attributes, from LinkedIn profiles, using the founder-specific LinkedIn URLs provided in Crunchbase. We also construct proxies for the founders' prior occupational exposure to tasks that can be performed with GenAI. We can match LinkedIn profiles for founders of approximately one-third of the startups in the Crunchbase sample; Panel D of Table 1 summarizes founder characteristics for this subsample.

3 A Forward-Looking Measure of GenAI Exposure

The extent to which startups can use GenAI depends on the nature of the tasks performed in their industry. Existing proxies for task-level GenAI exposure infer exposure from the similarity of job descriptions to textual descriptions of patents involving AI (Eloundou et al., 2024; Hampole et al., 2025), from expert assessments (Felten et al., 2019; Brynjolfsson et al., 2018; Shao et al., 2025), or surveys of workers (Humlum and Vestergaard, 2024). These measures describe what AI can do against a snapshot of today's task organization. They are informative about current capability, but less so about how the technology is diffusing.

Since our objective is to capture the effects of GenAI on startups that are early adopters of the new technology, we use the Anthropic Economic Index (AEI), which measures task reliance on GenAI from the realized choices of Claude.ai users. The AEI is constructed through a privacy-preserving analysis of millions of real-world conversations on Claude.ai. It measures the share of total queries involving a specific task among all queries to Claude during a week, including API queries. The index, therefore, captures the extent to which a task is performed by AI, either through automation or by augmenting human capabilities. We use the March 2025 version of the AEI, which is based on conversations during December 2024. Table A.1 provides examples of tasks with different exposure to GenAI. Because Claude.ai users are drawn disproportionately from early adopters of GenAI, the AEI serves as a forward-looking indicator of how the technology will be deployed as it diffuses more broadly. By capturing the frontier of GenAI adoption, the AEI is particularly well-suited to capture the effects on startups that are early adopters of the new technology.

Having obtained a measure of task exposure to AI, we follow the methodology pioneered by Autor et al. (2003) to obtain occupation and industry level proxies for AI exposure.

Specifically, we first aggregate approximately 19,000 distinct task-level exposures into the 923 occupations defined by the Bureau of Census O*NET occupational classification system. This gives us the AEI of each occupation based on the fraction of its constituent tasks that can be automated or enhanced by GenAI. As expected, computer programmers and IT-related jobs appear to be highly affected. Table A.2 provides more examples of occupations with high, average, and low exposure to GenAI.

Then, we map occupations to industries using the Bureau of Labor Statistics (BLS) Industry–Occupation Employment Matrix, which provides the employment share of each occupation within each industry. Formally, for each four-digit NAICS industry n, we compute:

$$AEI_n^{\text{NAICS4}} = \sum_o s_{on} \times AEI_o, \tag{1}$$

where AEI_o denotes the Anthropic exposure score for occupation o, and s_{on} represents the share of occupation o's employment in industry n. This approach yields a forward-looking index that captures each industry's exposure to generative AI technologies based on the composition of its occupational tasks.

We compare our AEI_n^{NAICS4} index with two other widely used indices of AI exposure. Across the 183 industries for which all measures are available, the AEI_n^{NAICS4} rank correlation with Felten et al. (2023) is 0.7 and with Eisfeldt et al. (2024); Eloundou et al. (2024) is 0.8, comparable to the rank correlation of 0.77 between the two benchmarks themselves. The AEI_n^{NAICS4} therefore identifies broadly the same ordering of industries as these alternative measures. This reassures us that our results are not driven by idiosyncrasies of the Claude.ai user base.

Since our other data sources provide different industry classifications or levels of granularity, we further aggregate the four-digit NAICS industry level measure to the two-digit sector level by taking employment-share-weighted averages across four-digit industries within each sector.

In the tests involving Crunchbase, which uses a proprietary industry classification, we need to adapt our proxies for industry exposure to GenAI. Crunchbase assigns each startup to one or more category groups (e.g., "Artificial Intelligence," "Financial Services," "Healthcare"), which overlap and are not mutually exclusive. [3] We construct a quasi-industry measure of GenAI exposure based on the startups' pre-GenAI occupational composition, as observed in Revelio Lab, within a Crunchbase industry category group. Let $AEI_{f,\tau}$ denote firm f's time-varying exposure in year τ, aggregated from Anthropic's occupation-level AEI using occupation weights within a firm. For each Crunchbase category group c, we construct an exposure measure that averages firm level AEI across all firms j assigned to that category during the pre-period (2020–2022):

$$AEI_f^{\text{pre}} = \frac{1}{|\mathcal{J}_c|\,|\mathcal{T}_{\text{pre}}|} \sum_{j \in \mathcal{J}_c} \sum_{\tau \in \mathcal{T}_{\text{pre}}} AEI_{j\tau}, \quad \mathcal{T}_{\text{pre}} = \{2020, 2021, 2022\}. \tag{2}$$

This measure proxies for pre-GenAI technological exposure at the quasi-industry level, capturing firms' occupational composition and category-group affiliations before GenAI diffusion. Firms belonging to multiple category groups are assigned the simple average of the corresponding AEI_j^{pre} values, yielding a firm-level measure (AEI_f^{pre}) that accommodates multi-category affiliations while reflecting the average AEI intensity of the relevant category groups prior to the widespread diffusion of the new technology.

3.1 Differences in Exposure to GenAI

Table A.3 provides examples of four-digit industries with high, median, and low levels of GenAI exposure. Software and other IT industries have high GenAI exposure; manufacturing, such as automobiles, and certain services, such as healthcare, have medium GenAI exposure; while industries that involve human interaction and manual work (e.g., delivery services, mining) have low GenAI exposure.

[3] While the merged dataset with Revelio includes a standard industry identifier (a six-digit NAICS code), this information is missing for roughly half of the firms in the combined Crunchbase–LinkedIn sample. Moreover, NAICS codes are not ideal for classifying Crunchbase startups because they are designed for mature industries, and innovative startups are concentrated in a handful of NAICS industries. In contrast, Crunchbase's own industry group classification better captures emerging sectors and groups together firms engaged in comparable business activities.

Figure D.1 shows the distribution of GenAI exposure across the 4-digit NAICS industries and the startups in our different samples. In Panel A, over 50% of industries appear to have no exposure to GenAI. Exposure to GenAI varies significantly across the remaining industries. In the most exposed industries, over 10% of tasks can be performed using GenAI.

The distribution of GenAI exposure differs significantly in the merged Crunchbase-Revelio sample. The firms in this sample are predominantly in the IT sector, as shown in Table A.4, and all of them have tasks that can be performed through GenAI. In a significant proportion of startups, over 50% of tasks can be at least partially performed by GenAI, suggesting that GenAI can have a profound impact on startup characteristics. Table A.5 describes the top-ranked startups for GenAI exposure. For instance, SuperDuperAi is an AI platform for creating professional videos. Experio Labs, Inc. offers an AI-powered knowledge management platform for startups.

Given the high dispersion of the GenAI industry exposure proxy, our tests compare high- and low-exposure industries, defined as those with exposure above and below the relevant sample median. We provide evidence in Section **??** that our results are robust to the measurement of task exposure to GenAI by using conversations with an alternative, widely used AI platform.

4 GenAI and New Business Formation

In this section, we first document the effect of GenAI on entrepreneurial entry at high frequency and with a coarse industry classification using BFS, and then move to granular industry-level data that allow us to characterize firm entry within narrowly defined industries. Finally, we examine the channels favoring entry by classifying new entrants into two non-mutually exclusive categories: *AI-native* startups, defined as firms whose products or services depend on GenAI, and *AI-enabled* startups, defined as firms whose operations rely on GenAI but whose output does not. We show that GenAI has favored the entry of both types of firms, suggesting the market-creation and cost-reduction mechanisms coexist.

4.1 High-Frequency Evidence

We use the Business Formation Statistics (BFS), which track new business applications in near real time, to test whether industries with greater GenAI task exposure experienced larger increases in entrepreneurial entry following its diffusion. We estimate the following difference-in-differences specification:

$$Y_{nmt} = \alpha + \beta\big(AEI_n^{\text{NAICS2}} \times Post_t\big) + \delta_n + \delta_{mt} + \varepsilon_{nmt}, \tag{3}$$

where the outcome variable, Y_{nmt}, is the number of projected new businesses in the two-digit NAICS industry n during month m of year t.

We define the post-period indicator ($Post_t$) as equal to one for months after August 2022, corresponding to the public release of *Stable Diffusion*, one of the first widely accessible text-to-image GenAI models. This timing coincides with the broader wave of mass-adopted foundation models—such as Midjourney's open beta in July 2022 and *ChatGPT*'s public release in November 2022, a period characterized by the improvement and rapid diffusion of foundation models and the widespread integration of generative tools into product development. It allows us to capture how the dramatic improvement in the new technology during the fall of 2022 affected entrepreneurial activity.

The variable AEI_n^{NAICS2} denotes the two-digit NAICS industry-level AI exposure index, aggregated from occupation-level AEI scores. Since aggregating the AEI to the industry level yields highly dispersed estimates across sectors, which may reflect both technological heterogeneity and aggregation bias, we classify sectors into two groups, above and below the sample median AEI, to enhance interpretability.

Estimates are obtained, including industry (δ_n) and month-year (δ_{mt}) fixed effects. Our sample period ranges from January 2020 to April 2025. Because the analysis is conducted at the two-digit NAICS sector level, the number of clusters is small (approximately 20 sectors), which prevents us from accounting for industry-level error correlations using standard clustering methodologies. For this reason, we compute standard errors using a bootstrap procedure with 500 repetitions, resampling at the sector and year-month level, to account for potential dependence among observations within the same industry and month.

Our empirical model enables us to trace the dynamic evolution of new business formation in relation to the diffusion of GenAI, comparing more GenAI-exposed industries to less-exposed ones while controlling for common time shocks and industry-specific effects. The coefficient β measures the differential change in projected business formation across industries with varying levels of pre-existing GenAI exposure following the diffusion of GenAI.

Table 2 reveals strong positive effects of GenAI diffusion on new business formation in industries relatively more exposed to GenAI. We consider different measures of new business formation, including: (i) All business applications (column 1); (ii) Applications from businesses that are estimated to have a high propensity of becoming employers, due to the presence of planned payroll in their applications (column 2), which we refer to as *High-Propensity Business Applications* (HBA); and (iii) Corporations' business applications (column 3). The estimated effects are not only statistically significant but also economically meaningful: In industries with GenAI exposure above the median, we observe a 3.8% larger increase in monthly business applications (column 1) and in business applications estimated to have a high probability of becoming employers (column 2). The estimated effect is even larger for corporate business applications, which experience a 5.5% higher monthly increase in industries with GenAI exposure above the median following the diffusion of GenAI. We also observe substantial positive effects of the diffusion of the new technology in industries with high GenAI exposure for *Projected Business Formations* (PBF), which are the Census Bureau's projections of business propensities to become employers within four to eight quarters columns 4 and 5). This evidence suggests that GenAI is significantly impacting entrepreneurial activity.

To address concerns that we may capture pre-existing trends, we examine the dynamic patterns in Figure 1. The positive effects of GenAI diffusion begin to emerge in late 2022, shortly after the release of early foundation models, and intensify over time as the new technology diffuses and improves. There is no evidence of systematic pre-trends prior to this point, supporting the validity of the difference-in-differences identification.

4.2 Firm-Level Entry and Industry Composition

The BFS data are well-suited to detecting the timing of entry but two-digit NAICS industry coverage is coarse. We therefore turn to firm-level data from InfoGroup, which allows us to measure entry at the four-digit NAICS level.

We consider newly established, single-location firms, treating each firm's founding year as the observation year. For each year between 2020 and 2024, we count newly founded firms at the four-digit NAICS level. To capture the differential change in firm entry across industries with varying degrees of pre-existing GenAI exposure following the diffusion of such technologies, we estimate a difference-in-differences specification similar to the one in the previous section but with all industry variables defined at the four-digit level. Accordingly, we include industry fixed effects and cluster standard errors at the four-digit NAICS level.

Table 3 reports the estimates. We compare industries with exposure to GenAI above/below the median, as in the earlier analyses. We find that in more AI-exposed industries, the probability of observing any entry (an indicator equal to one if at least one new firm is established in an industry-year cell) increases after the diffusion of GenAI (columns 1 and 2). Column 2 of Table 3 implies that industries with GenAI exposure above the median have a 10% higher probability of any new firm entering during a year following the diffusion of GenAI. This result suggests that the likelihood of new firm creation has increased in industries that were already technologically proximate to GenAI before 2023. Moreover, industries with greater pre-existing AI exposure have more new firms entering post-GenAI (columns 3 and 4 of Table 3). Results are again not only statistically but also economically significant. Column 4 of Table 3 implies that new firms increase by over 20% more in industries with GenAI exposure above the median following the introduction of GenAI, relative to less-exposed industries. Overall, it appears that GenAI lowered the fixed costs of entry or opened new market opportunities in these sectors.

We also explore the sensitivity of our estimates when considering different types of interactions with Claude. Claude.ai data enables us to examine the types of interactions with GenAI to perform different tasks. Specifically, we distinguish between automation and augmentation modes of using Claude.ai, based on the fraction of interactions in which the AI directly executes tasks with minimal human involvement (automation) and those in which AI and humans collaborate (augmentation). We multiply the frequency of interactions for each task by the fraction of interactions in a given mode to obtain the automation- and augmentation-oriented interaction frequencies. We then aggregate at the task and occupation levels. Figure 2 presents the estimates using industry measures of automation-oriented and augmentation-oriented GenAI exposures. The estimates are qualitatively and quantitatively similar to the baseline, suggesting that task exposure through both modes facilitates entrepreneurial activity in new firms.

We also exploit the fact that Claude.ai further decomposes the automation mode of interaction into two distinct categories: conversations that are directive or involve a feedback loop. Furthermore, augmentation mode interactions are decomposed into three categories: conversations that involve users' learning, validation of users' work, or task iteration, that is, conversations in which users interact collaboratively with Claude.ai. We apply the same process as for the automation and augmentation modes to investigate the effects of exposure to GenAI across these more specific modes of operation. Again, we observe that an industry's exposure to tasks involving automation in directive modes or learning, user-performed work validation, or iterative collaboration between the user and Claude.ai positively affects new firm formation, consistent with the baseline estimates. We do not observe a statistically significant effect for feedback loop interactions, defined as interactions in which AI is guided by environmental feedback rather than by a human. This mode may be too advanced at the current stage of the GenAI technology's diffusion.

4.3 How Does GenAI Favor Entry?

The evidence we have presented so far is consistent with the hypothesis that GenAI favors entrepreneurial entry, but it is silent on the mechanisms driving its effect. GenAI can increase entry through two non-mutually exclusive channels. First, GenAI may enable startups to build new products and services around generative capabilities, creating investment opportunities that did not previously exist. Second, by relying on pay-per-use platforms offered by foundation model providers such as OpenAI, Anthropic, and Google DeepMind, or on other GenAI applications, startups may automate internal operations and face lower entry costs even when GenAI is not central to their value proposition.

To capture these two mechanisms, we distinguish between *AI-enabled* startups, defined as firms that use GenAI to automate internal processes, and *AI-native* startups, defined as firms whose products or services depend on GenAI. The two categories are not mutually exclusive: a firm that both sells AI products and uses AI internally is classified under both labels.

Classification methodology. Identifying these two categories is challenging because Crunchbase profiles describe what a firm sells but rarely document how it operates internally, and self-reported labels are noisy. We therefore construct two complementary classification pipelines based on verifiable external evidence. We describe both pipelines in detail in Appendix B and summarize their structure here.

The first pipeline identifies AI-enabled startups through a two-stage agentic procedure. In the first stage, we assess whether a startup is sufficiently searchable online by combining a website health check with a query to the Tavily search API; firms with weak signals are not escalated. In the second stage, escalated firms are processed using Perplexity's deep-research API, which performs iterative search-and-reason steps to retrieve and validate evidence across multiple online sources. The pipeline is configured to enforce strict evidence standards: every positive finding must be tied to a citable URL, the agent reports only what sources explicitly state, rather than inferring adoption from ambiguous language, and a separate guardrail distinguishes selling AI products from using AI internally. A startup is classified as AI-enabled if the resulting dossier contains verifiable evidence that the firm uses GenAI tools in at least one internal function (e.g., engineering, marketing, operations, customer service, sales, human resources, or finance). [4]

The second pipeline identifies AI-native startups by combining Crunchbase information with website evidence and Tavily-retrieved crawl text, and mapping each firm into a taxonomy that distinguishes foundation-layer firms, applied vertical AI firms, AI tooling firms, autonomous-agent firms, generative content platforms, thick LLM integrators, and thin LLM wrappers. Firms outside the taxonomy are classified as traditional tech, non-tech, or AI-adjacent. For each AI-native firm, the pipeline also returns a measure of reliance on third-party AI APIs (high, medium, or low), which allows us to separate firms whose core product depends structurally on external models from those that combine external services with substantial proprietary infrastructure.

Coverage and validation. Among approximately 260,000 startups in Crunchbase, 44,000 have enough information to enter our pipeline. Among these, 21% pass the AI-enabled screening and are escalated to deep research; and, 13% yield at least one verified adoption finding. The named-tool distribution is informative about the use of GenAI tools: ChatGPT and other OpenAI products account for 49% of mentions, Claude and other Anthropic products for 22%, GitHub Copilot for 12%, Cursor for 11%, and Gemini and other Google AI products for 6%. The supporting sources are company blogs (36%) and job postings (25%), with the remainder drawn from podcasts, interviews, and press coverage. Importantly, the fact that AI-enabled startups name a broad range of tools reassures us that our AEI-exposure measure, though constructed from Claude.ai conversations, captures a technology whose adoption patterns extend well beyond any single platform. For the AI-native

[4] For example, the pipeline identifies Sawmills as AI-enabled in engineering based on a company blog in which the CTO describes adopting Claude Opus, GitHub Copilot, and GPT-based models for planning, execution, and debugging in software development; Infinite Renewals as AI-enabled in marketing based on a podcast in which the founder explicitly states using Jasper for content creation; and Daisy as AI-enabled in customer service based on a blog post describing how technicians use ChatGPT to assist with troubleshooting and customer queries.

pipeline, 48% of startups have sufficient website evidence to be classified, and 15.6% of these are identified as AI-native. The AI-native rate is substantially higher among firms founded in the GenAI era, reaching 42.8% for startups founded in 2023 or later.

Diffusion patterns. Figure 3 shows that the shares of both AI-enabled and AI-native startups increase sharply after 2022. The rise is steeper for AI-native startups, consistent with a process of technological diffusion that has led to the emergence of new products and services built around generative capabilities. The share of AI-enabled startups is lower than that of AI-native startups in absolute terms, but this gap likely understates internal adoption: AI-enabled use is harder to identify because startups are less likely to advertise their internal tool stack than their product offering. The combined evidence suggests that the surge in entrepreneurial entry is not driven solely by reliance on foundation models to automate operations and reduce entry costs; a substantial proportion of new firms are creating GenAI-based products and services in a way that can be conducive to exponential growth.

The differences in entry rates we have documented so far are driven by both types of startups. Figure 4 plots the share of AI-enabled startups in each Crunchbase industry category group against the group's GenAI exposure, separately for each year before and after the new technology's diffusion.[5] The relation is positive in every year, but it becomes appreciably steeper in 2023 and 2024 relative to 2020 and 2021. Figure 5 shows the same pattern for AI-native startups: the slope of the relation between the share of AI-native firms and category-level GenAI exposure steepens markedly after 2022, suggesting that technological diffusion has led many startups to develop AI products.

In sum, evidence in Figures 4 and 5 demonstrates that entry into GenAI-exposed industries is indeed driven by businesses that rely on GenAI, rather than by industry-level factors unrelated to the new technology. More importantly, these tests provide direct evidence on the mechanisms through which GenAI favors business dynamism. While GenAI is favoring

[5]We omit 2022, the intermediate year.

entry by enabling startups to automate tasks and lowering entry costs, it is also enhancing their AI product offerings, with the potential to increase economic growth.

We also provide evidence on how startups are incorporating AI. Figure 6 reports the functions in which AI-enabled startups most frequently deploy GenAI applications. The most common use cases are engineering, marketing, and operations, followed by product and design, sales, customer service, human resources, and finance and accounting. All these functions involve tasks with high exposure to GenAI in our AEI measure, providing a consistency check between the deep-research classification and the task-based exposure index used in our main specifications.

AI-enabled startups naturally rely on third-party foundational models. It is also interesting to know to what extent AI-native firms are developing new models. We find that foundation-layer firms, which build core models, infrastructure, or data layers, account for just 1.3% of AI-native startups. The overwhelming majority instead builds products and services on top of third-party foundation models. Applied vertical AI firms, which embed generative capabilities into products for specific industries or workflows, are the largest category at 57.0% of AI-native startups, followed by thick LLM integrators (19.1%), AI-native tooling firms (10.0%), and autonomous-agent firms (9.6%). Generative content platforms (1.8%) and thin LLM wrappers (1.2%) are comparatively rare. The RAD score, which measures the extent to which an AI-native startup depends on third-party AI APIs for its core product functionality, confirms that reliance on external models is pervasive. We classify 43.8% of AI-native startups as RAD-H, indicating high structural reliance on external AI models or APIs, and a further 51.0% as RAD-M, indicating meaningful use of external AI services combined with proprietary workflow, data, or product layers. Only 5.2% are classified as RAD-L, indicating relatively low dependence on third-party AI APIs. Thus, although a large share of AI-native entrants add nontrivial proprietary layers on top of external infrastructure, almost none operate independently of it. Overall, these patterns indicate that GenAI-era entrepreneurship is built predominantly on foundation models supplied by a small number of incumbents rather than on models developed in-house, so that the core capabilities of most AI-native entrants depend on continued access to third-party APIs.

5 GenAI and Changes in Industry Structure

5.1 Geographic Dispersion of New Entrants

Human capital externalities and financial frictions generate agglomeration economies and tend to concentrate entrepreneurial activity in locations that offer thick labor markets and investor networks (Ellison and Glaeser, 1997; Ewens and Farre-Mensa, 2022). If GenAI provides on-demand expertise and weakens human capital complementarities, founders may no longer need to be physically close to specialized human capital. Similarly, if GenAI reduces the need to scale the labor force for acquiring market share, new ventures' financing needs may decrease. By automating processes and augmenting founders' capabilities, GenAI may thus decrease entry costs the most in locations that previously lacked human capital externalities and privileged access to entrepreneurial finance. Evidence that entrants concentrate even more, by contrast, would suggest complementarity between GenAI and agglomeration, perhaps because access to frontier talent, compute partnerships, or proprietary data becomes more valuable in an AI-native world.

InfoGroup data allows us to explore how GenAI is affecting the geographic distribution of new ventures because it provides firms' locations. Table 4 asks whether the number of geographic areas with entry in a given industry during a year increases post-GenAI and estimates the difference-in-differences model used in the previous tests to evaluate any changes. In addition to the number of geographic areas, we measure the concentration of new firms in a given industry and year using a geographic Herfindahl–Hirschman Index (HHI), defined as the sum of the squares of the shares of new entrants across geographic locations for that industry and year.

We find that firm entry becomes more geographically dispersed following the diffusion of GenAI. Across specifications, we increase the granularity of geographic areas, progressing from states (columns 1 and 4) to counties (columns 3 and 6) via combined statistical areas

(CSAs) in columns 2 and 5. More states, CSAs, and counties record new firm formation in highly exposed industries after 2022 (columns 1 to 3), and the HHI of firm creation declines correspondingly (columns 4 to 6). The effects are not only statistically significant but also economically meaningful. For instance, the coefficient estimate in column 6 implies an over 16% more pronounced decrease in the county level concentration of entry in industries more exposed to GenAI. These patterns suggest reduced entry barriers: GenAI may reduce agglomeration economies and enable entrepreneurs outside traditional innovation hubs to enter AI-adjacent markets.

Table 5 investigates the mechanisms driving the increasing geographic dispersion of entrepreneurial activity. GenAI may substitute for scarce, specialized human capital, thereby reducing the need for local knowledge. Since startups may not need to scale their workforce to increase market share, they may also have lower funding needs and thus benefit less from locating in areas with historically concentrated VC investment. We measure geographic industry specialization by assessing the pre-GenAI distribution of local employment. For each county and four-digit NAICS industry, we compute the share of workers employed in that industry relative to total county employment, using the 2020-2022 pre-treatment average. Counties in the top decile of this distribution are classified as having higher availability of specialized labor and knowledge in the respective industry.[6] We then compare post-GenAI entry responses across areas with high versus low specialization. In Panel A of Table 5, we find that the rise in firm formation in GenAI-exposed industries is concentrated in geographic areas characterized by lower labor specialization in the entrant's industry (columns 1 and 3). Thus, GenAI appears to reduce the need for industry-specific human capital and the benefits arising from local knowledge spillovers.

To examine the VC financing channel, we exploit the spatial concentration of venture capital activity and test whether GenAI-related entry expands more strongly in traditionally

[6] Results are robust to alternative thresholds, including the top 5th percentile and top quartile of the specialization distribution. They are also robust to defining specialization at the CSA rather than the county level; see Panel A of Table D.2.

underserved areas. We classify localities as VC-hub or non-hub states based on pre-GenAI venture-finance intensity. California, Massachusetts, and New York, which have the highest levels of VC activity, are identified as VC hubs. We then compare post-GenAI entry dynamics between VC-rich and VC-scarce areas. The results in Panel B of Table 5 show that the increase in new firm formation in GenAI-exposed industries is disproportionately larger in states with lower VC presence, suggesting that GenAI reduces entrepreneurs' dependence on local venture capital ecosystems and enables firm creation in regions previously constrained by financing access. We reach the same conclusions if we define venture capital hubs at the county level by aggregating the total dollar value of investments made by U.S.-based VC investors in each county using Crunchbase data and ranking counties by this measure. Counties in the top ten of this distribution are designated as top VC counties. Using this alternative definition, we show (Panel B of Table D.2) stronger GenAI-induced entry in counties with limited VC activity.

5.2 The Size and Organization of New Entrants

If GenAI lowers entry costs by reducing human capital complementarities, we should observe that new entrants are smaller, as the minimum efficient scale of human capital declines. Furthermore, weaker skill complementarities should reduce the number of distinct occupational roles that startups need to staff. To test whether newly founded firms operate with fewer employees and a narrower set of occupations, we examine these two dimensions with different datasets, beginning with the size distribution in InfoGroup and turning to workforce composition in the merged Crunchbase–Revelio sample.

Using InfoGroup data, we count the number of new firms in an industry and year below a size threshold to test whether newly founded firms have become smaller in the post-GenAI period. We present the results in Table 6. We find that newly created firms in more exposed industries tend to be smaller. The increase in entry is concentrated among firms with initial employment below the pre-treatment industry median (columns 1 and 3), while the probability of entry for firms above the median industry size and the number of firms above the median size remain unchanged (columns 2 and 4). Using absolute size thresholds (e.g., 3/5/10 employees) produces similar results (Table D.3). This pattern suggests that GenAI diffusion is associated with a shift in composition toward smaller, leaner startups. It thus appears that access to GenAI tools enables entrepreneurs to substitute technology for labor in the early stages of firm formation, reducing the need for large founding teams.

Such an interpretation would imply that GenAI-exposed startups also reduce the mix of occupations they bring in at the time of founding, as some functions can be automated or augmented by GenAI. To test this, we turn to the merged Crunchbase–Revelio sample, which provides employee-level occupational data for a subset of entrants. This allows us to test whether the decline in size reflects a proportional contraction of the founding team or a reallocation across occupations consistent with GenAI substituting for some tasks and complementing others.

To examine how startups' characteristics evolve in industries differentially exposed to GenAI, we estimate the following empirical model:

$$Y_{ft} = \alpha + \beta\big(AEI_f^{\text{pre}} \times Post_t\big) + \delta_t + \sum_{c\in\mathcal{C}} \phi_c D_{fc} + \varepsilon_{ft}, \tag{4}$$

where the dependent variable, Y_{ft}, is a firm-level outcome, capturing characteristics of the firm's founders and employees, measured at the end of the founding year. Thus, the unit of observation is a startup firm f founded in year $t \in \{2020, \ldots, 2024\}$. We consider the following three outcomes: the total number of employees by the end of the founding year (*# employee*), the number of unique occupations among the firm's employees by the end of the founding year (*# occupation*), and the Herfindahl–Hirschman Index of occupational concentration within the firm, capturing within-firm task diversity (*HHI occupation*).

The variable AEI_f^{pre} is the firm's ex-ante GenAI exposure, defined in Section 3, based on its Crunchbase category groups. As in the earlier tests, the indicator $Post_t$ takes the value one for the post-GenAI period ($t \in \{2023, 2024\}$) and zero for the pre-GenAI period ($t \in \{2020, 2021, 2022\}$). Thus, coefficient β captures how the composition of a startup's founding workforce changes in the post-period as a function of its exposure to GenAI. As in the earlier tests, we compare startups with exposure above the median with those with exposure below the median. We include founding-year fixed effects (δ_t) and industry category indicators (D_{fc}) that take the value 1 if firm f is associated with category group c. Because firms in Crunchbase can belong to multiple industry category groups, several of these dummy variables may be equal to one simultaneously. Since industry categories are not mutually exclusive, we use bootstrapping to estimate standard errors.

Table 7 confirms that, relative to less-exposed peers, startups with higher ex-ante GenAI exposure founded in 2023–2024 have *fewer employees*. More importantly, we find that these startups have *fewer distinct occupations* and *higher occupational concentration (HHI)* among their employees at the end of the founding year. The coefficient estimates are largely invariant to whether we include founding-year fixed effects. The estimates are not only statistically significant but also economically meaningful. The average startup in our sample has 4.6 employees. Thus, the estimate in column 1 implies a nearly 25% decrease in the number of employees for startups with GenAI exposure above the median following the diffusion of GenAI. In column 3, the concentration of occupations in startups with GenAI exposure above the median increases by nearly 10% following the diffusion of GenAI.[7]

Figure 7 provides a visual characterization of the effects of GenAI exposure on startup size. We plot the frequency of startups by employee count during the pre- and post-periods, distinguishing between high-GenAI-exposure industries (Panel a) and low-GenAI-exposure industries (Panel b). The frequency of startups with only one employee increases significantly in the post-period in high-GenAI industry categories. This pattern is not observed in low-GenAI-exposure industries, which experience an increase in the number of startups with more than 10 employees. This evidence confirms that sample selection is unlikely to drive our findings, since small startups are more likely to be underreported in the merged Crunchbase-

[7]All of the above results remain robust when using four-digit NAICS codes with corresponding industry fixed effects and clustering at the NAICS level instead of using Crunchbase industry category indicators. However, the sample size is reduced by roughly half, since about 50% of the merged firms lack NAICS assignments in Revelio.

Revelio sample in recent years.

Taken together, these results suggest that firms with greater GenAI exposure start with smaller and more specialized teams. This pattern is consistent with the notion of "lean entrepreneurship," characterized by a narrower scope of roles and a concentration of human capital within founding teams. It suggests that by reducing cognitive costs, GenAI decreases skill complementarities and decreases the demand for labor and large founding teams.

6 GenAI and Labor Demand of Startups

6.1 Aggregate Employment Effects in New Entrants

Our findings indicate that GenAI has increased firm entry while reducing startups' employee headcount. From an aggregate perspective, it is important to understand whether new firm formation mitigates concerns about widespread labor displacement and depressed labor demand, or instead accentuates them due to startups' smaller size.

To evaluate these aggregate effects, we examine how the net impact of these two forces shapes job creation in startups relatively more exposed to GenAI. To do so, we use QWI, a nationally representative administrative measure of employment and earnings for very young firms (age 0 and 1) and other companies at the NAICS-4 quarter level. Applying the same difference-in-differences framework as in the previous sections, we compare employment and wages in more versus less-exposed industries before and after the diffusion of GenAI. Table 8 shows that employment among the new firms rises by 7.4% in more exposed industries (column 1), while employment among older firms shows no significant change (column 2). We also test whether monthly earnings in industries more exposed to GenAI have increased, as this would suggest increased labor demand in newly started firms in these industries. Consistent with the conjecture that demand for labor of startups in industries exposed to GenAI increases, we find that average monthly earnings increase by about 5% for workers in very young firms in high-exposure industries in comparison to the average industry (column 3), whereas monthly earnings for workers in high-GenAI-exposure incumbent firms remain flat (column 4).

The lack of change among established firms likely reflects incumbents' slow adoption of technology and does not dispel concerns about job displacement. However, the findings that aggregate employment at newly established firms increases, notwithstanding the smaller size of startups, reveal an early countervailing channel not emphasized in prior work: the surge in new entrants generates net employment expansion and higher wages among the youngest firms in industries most affected by GenAI, even though these firms are started with fewer employees than in the pre-GenAI period.

6.2 GenAI and Startups' Occupations

To the extent that increases in labor demand are likely to come from startups, it is important to understand the skills they require. We therefore explore the occupational composition of startups and how it evolves with the diffusion of GenAI.

We begin by examining how the jobs created by startups vary by occupational exposure to GenAI. Occupations are grouped into three categories based on their AEI values: *high* exposure ($AEI > 0.1$), *medium* exposure ($0.01 \leq AEI \leq 0.1$), and *low* exposure ($AEI < 0.01$). For each startup, we count the number of employees in each group as of the end of the founding year and use these counts as dependent variables in the same difference-in-differences framework used in the earlier tests. Table 9 reports the results. Employment in high-AEI occupations remains roughly stable or slightly increases, while employment in medium- and low-AEI occupations declines significantly, with the largest relative reduction occurring in low-AEI roles. These differences should be understood as relative adjustments across industries with different levels of exposure to GenAI, rather than as absolute changes in startups' occupational composition.

A natural question is whether this reallocation is concentrated in startups that actually use or build on GenAI. To address this question, we exploit the firm-level classifications developed in Subsection 4.3 and compare the founding-year occupational composition of startups that verifiably use GenAI internally or sell AI-based products to that of otherwise similar startups founded in the same period. Figure 8 reports the average number of employees per startup at the end of the founding year, decomposed into low-, mid-, and high-AEI occupational groups, separately for AI-enabled and non-AI-enabled startups (Panel a) and for AI-native and non-AI-native startups (Panel b). The patterns reveal both a contraction in entry size and a reallocation toward GenAI-intensive occupations among the cohorts that most directly benefit from the improvement and diffusion of the new technology. AI-enabled startups founded before 2022 enter with 7.56 employees on average, whereas those founded after 2022 enter with 5.35; over the same period, non-AI-enabled startups enter with around 4-5 employees in both periods. Among AI-enabled entrants, the share of high-AEI employees at founding rises from 15.6% to 19.8%, the mid-AEI share rises from 15.8% to 18.5%, and the low-AEI share falls from 68.7% to 61.7%. Non-AI-enabled cohorts show similar but weaker changes, likely because the label is subject to false negatives: our classification requires verifiable external evidence of internal GenAI use, which many adopters may lack.

The contrast is sharper in Panel b. Average entry size among AI-native startups declines from 4.69 to 3.87 employees, the high-AEI share rises from 20.3% to 25.4%, the mid-AEI share rises from 16.6% to 18.3%, and the low-AEI share falls from 63.1% to 56.3%, while non-AI-native cohorts show essentially no change in either entry size or composition. Overall, GenAI appears to be reshaping the organization of new ventures, with the firms most directly exposed to the technology entering both leaner and more concentrated in GenAI-exposed roles than otherwise comparable firms founded in the same period.

We next ask which specific occupations drive these patterns. Figure 9 shows the occupations that experienced the largest employment losses (Panel a) and gains (Panel b) at startups in the post-GenAI period relative to the earlier period. Software developers recorded the largest occupational gains, rising from a 6.6% employment share pre-GenAI to 11.3% following the diffusion of the new technology. Similar occupations, such as engineers and information technology project managers, which are relevant to the adoption of the new technology, also experience occupational gains in startups. The largest decreases in employment shares are among chief executives, marketing managers, and general operations managers, roles that involve coordination and may have been replaced by GenAI or are less relevant in smaller companies. These occupations are, on average, significantly less exposed to GenAI but appear to remain in higher demand at established companies (Hampole et al., 2025). Figure 10 confirms that the occupations gaining employment at startups have higher GenAI exposure. The positive relation between an occupation's AEI and its change in startup employment share holds both for the full sample of occupations (Panel a) and for those in the top decile of the pre-period startup employment share (Panel b), indicating that the pattern is not driven by small, marginal occupations.

Taken together, these results suggest that GenAI reduces the minimum efficient scale of human labor at founding. This raises within-firm occupational concentration and the average GenAI exposure of the workforce, while reducing headcount and the variety of in-house occupations. These findings are consistent with evidence from time-use surveys, which show that working hours in occupations using GenAI have increased following the diffusion of the technology, presumably due to increased productivity (Jiang et al., 2025). Low-AEI occupations, by contrast, are less complementary to GenAI-related technologies, leading startups in more-exposed categories to scale back hiring in these roles.

Importantly, the occupational composition of startups differs sharply from that of mature companies. While high-AEI occupations appear complementary to the new technology in startups, established companies have been found to cut jobs in occupations most exposed to GenAI, such as software development (Acemoglu et al., 2022; Brynjolfsson et al., 2018). Demand for occupations with high GenAI exposure in startups thus appears fundamentally different from that in large, mature companies. The investment opportunities created by the new technology appear to necessitate domain knowledge and increase the demand for high-AEI roles in startups, mitigating concerns about widespread job displacement.

6.3 Founder Characteristics

We also examine whether founders' skills, prior experience, and other demographic characteristics vary across industries with different exposure to GenAI. We estimate a regression similar to Equation 4, using startup-founder-level observations constructed from snapshots of startups founded between 2020 and 2024.

Given our results on the jobs created by startups, we first check whether founders were exposed to GenAI-related tasks in their pre-founding careers. We consider founders' previous occupations and their exposure to GenAI, constructed using Anthropic's occupation-level AEI. Specifically, we examine the GenAI exposure of the founder's last occupation before founding ($PriorAEI_{kf}^{\text{most recent}}$); the average GenAI exposure across all the founder's pre-founding occupations ($PriorAEI_{kf}^{\text{mean}}$); and the maximum GenAI exposure observed across all pre-founding occupations ($PriorAEI_{kf}^{\text{max}}$).

Table 10 shows that founders of startups in more GenAI-exposed categories have higher exposure to GenAI in their past occupations. This does not merely reflect the founder's tendency to start businesses in industries where they already have experience because the effect becomes even more pronounced following the diffusion of GenAI. This pattern implies that, after the diffusion of GenAI, startups in more exposed categories are increasingly founded by individuals whose past work involved tasks more closely associated with GenAI-exposed occupations, suggesting that domain knowledge may be important to exploit the new technology.

To provide further evidence on the skills that matter in startups, we move beyond structured measures of founders' prior work experience to examine how founders describe themselves at the time of entry. Whereas the analysis above focuses on founders' occupational exposure to GenAI-related tasks, which captures hard skills accumulated through past employment, Figure 11 provides complementary evidence of how the content of founders' self-described characteristics shifts following the diffusion of GenAI. Using an LLM-assisted pipeline described in Appendix C, we extract granular traits from founders' LinkedIn profiles around the founding year, organize them into a bottom-up taxonomy, and compare cluster shares before and after GenAI diffusion. The top panel focuses on traits explicitly stated in founders' profiles, while the bottom panel also includes traits inferred from their narratives and actions. Across both panels, we observe declines in traits associated with roles and activities such as operations, service delivery, logistics, and certain design-oriented tasks, alongside increases in traits associated with roles that may be interpreted as more complementary to GenAI, including AI/ML development, software engineering, and cloud infrastructure. Consistent with our earlier findings, the distribution of founder traits shifts toward more technical, task-specific ("hard") capabilities and away from broader, dispositional or values-oriented ("soft") attributes.

Table 11 confirms these patterns in a regression framework that controls for the total number of observed traits extracted from each founder's profile and for industry-category fixed effects. Among founders of startups in industries with above-median GenAI exposure, the count of traits in the AI/ML and GenAI cluster rises by 0.247 after 2022 (column 2), against a sample mean of 0.217, indicating that the share of founders describing themselves in terms of AI and machine learning capabilities roughly doubles in more-exposed industries following the diffusion of the new technology. By contrast, traits associated with capital raising and deal mechanics, product development, managerial and sales roles, and operations and logistics decline differentially in more-exposed industries, with the largest declines for capital and investing (-0.134, column 1) and operations and logistics. These shifts are robust to the inclusion of industry-category indicators and founding-year fixed effects (columns 6–10), suggesting that they reflect a recomposition of the founder pool rather than secular trends in how founders in specific industries describe themselves on professional platforms. The patterns are also robust to extending the trait counts to include traits logically inferred from founders' narratives and actions, as reported in Table D.4 in the Internet Appendix, indicating that the shifts we document are not driven by changes in narrative style.

The shifts in founders' self-described capabilities also help rationalize the increase in entry we observe in regions outside traditional venture capital hubs we documented in Section 5.1. Founders de-emphasizing the capital-raising and deal-making skills that have historically been central to venture-backed entrepreneurship are plausibly less reliant on proximity to venture capital networks.

Overall, in industries where GenAI is most relevant, the individuals who become entrepreneurs combine technical AI capabilities with domain knowledge from past occupations to develop new ventures. It is plausible that individuals with prior experience in AI-intensive occupations may be better equipped to recognize and capitalize on opportunities created by GenAI technologies. However, these same occupations are among the most affected by technological disruption in established companies, potentially prompting some individuals to leave traditional employment and pursue entrepreneurship as a response.

To shed some light on the drivers of entry, Table 12 examines how founders' professional, educational and demographic characteristics evolve differentially in startups with above-median GenAI exposure. We start by analyzing the characteristics of the founders' previous employers. We conjecture that if founders previously worked at firms that were expanding their employment, the founders are unlikely to have been pushed into entrepreneurship because they lost their jobs. In column 1, the dependent variable is an indicator equal to one if the founder's previous employer belongs to the top decile of the employment-growth distribution over the 12-month window surrounding the founder's job transition. We observe that founders are more likely to come from these companies with fast-growing employment. In contrast, column 2 shows no corresponding increase in the likelihood that founders come from downsizing employers in the bottom decile of the employment-growth distribution. This is consistent with evidence that entrepreneurial firms spawn entrepreneurs (Nanda and Sørensen, 2010; Gompers et al., 2005) and does not support the notion that individuals are forced into entrepreneurship by downsizing at their previous employer.

We also consider the founder's position at their previous employer and their compensation to assess whether GenAI may be attracting lower-quality or less experienced individuals into entrepreneurship. In column 3, the dependent variable is a dummy that takes the value of one if the founder held a managerial position before starting the new business. We observe that founders of businesses more exposed to GenAI are less likely to have held managerial positions before starting their new company, suggesting they have lower seniority. We reach the same conclusion in column 4, where we rank the seniority of the position the founder held prior to founding the company. However, founders do not appear to have fewer years of experience since their first job (column 5), indicating that founders of businesses highly exposed to GenAI held technical positions that did not involve managerial and coordination tasks.

Not only is such an interpretation consistent with evidence from managerial traits, but it is also supported by the founder's compensation before founding. If founders did not hold managerial positions because they were less successful or of lower quality than founders in the same industries prior to the diffusion of GenAI, we would expect them to have had lower compensation before founding than their counterparts in industries with low exposure to GenAI. We do not find any changes in prior-employment compensation for founders of businesses highly exposed to GenAI following the diffusion of the new technology (column 6). This suggests that GenAI is unlikely to be drawing lower-quality individuals into entrepreneurship. Overall, the evidence is more consistent with opportunity-driven entry: founders are drawn into entrepreneurial activity by the investment opportunities the new technology creates, rather than pushed into it by displacement or weaker outside options. Such a conclusion is further reinforced by the fact that the founders do not become more likely to suffer a pay cut when they start a new firm (column 7).

Finally, we consider founders' educational attainment, coded on a six-point scale from high school to doctorate, and an indicator variable capturing whether the founder is male. Founders in more-exposed industries exhibit lower educational attainment after 2022, with a coefficient of -0.182 on the interaction term in column 8 against a sample mean of 3.93 on the six-point scale. The coefficient implies a modest but statistically meaningful shift away from graduate-level credentials in more-exposed industries. By contrast, we observe no differential change in the share of male founders (column 9). The lack of a differential shift in gender suggests that the persistent underrepresentation of women among founders is not solely driven by financial barriers.

Combined with the trait-level evidence in Table 11, these results suggest a compositional shift in the founder pool following the diffusion of GenAI toward individuals with domain knowledge in occupations highly exposed to GenAI and technical capabilities rather than formal academic credentials and managerial experience, but without a deterioration in the quality of the marginal entrant.

7 Robustness

7.1 Alternative Measures of Exposure to GenAI

As discussed in Section 3, the AEI relies on a specific platform, Claude.ai. To address this concern, we use an index, analogous to the AEI, constructed from conversations with *Microsoft Bing Copilot*, which is publicly available at `https://github.com/microsoft/working-with-ai`. Specifically, Microsoft provides the *AI Applicability Index*, which not only captures how frequently Copilot is used for a given work activity, but also measures how effectively Copilot completes or assists with that activity and the share of the occupation's work that the activity represents. Unlike the AEI, which measures the share of user conversations or traffic across O*NET tasks, Microsoft's index covers O*NET 330 intermediate work activities (IWAs). These can be aggregated to the occupation level using the occupational activity weights, based on O*NET intermediate work activities. At the O*NET occupation level (approximately 750 occupations), the rank correlation between the two raw measures is 0.65, indicating that the two indices capture a broadly consistent ordering of occupations by their exposure to GenAI. Once converted to the four-digit NAICS industry level, the rank correlation approaches 0.7, suggesting strong consistency in relative exposure patterns across industries. Figure D.2 provides graphical evidence on the comparability of the distribution of the two measures.

We re-estimate all our empirical models using the exposure measure constructed from the Copilot data and find qualitatively similar results: industries more exposed to GenAI experience higher entry rates, driven by smaller, leaner firms, and startups are more geographically dispersed with increased entry in areas that exhibit less labor specialization and less availability of venture capital funding. Figures 12 and 13 compare point estimates from the GenAI exposure measure based on Microsoft Bing Copilot with those from the AEI and show that the estimates are quantitatively and qualitatively similar.

7.2 Types of GenAI Interactions

Figures 12 and 13 also compare the estimates of the main specifications on new business formation and firm entry across different definitions of GenAI. First, Anthropic released a newer version of the index in September 2025. Our estimates are invariant if we define GenAI exposure using the September 2025 version of the AEI, which is based on August 2025 conversations, rather than the December 2024 conversations released in March 2025, as used in the baseline estimates.

Second, a possible concern is that AEI does not allow us to distinguish between business and personal use. However, prompts related to non-commercial activities, such as trip planning or entertainment, lack clear counterparts in O*NET and therefore cannot be matched to occupational tasks. Consequently, our task-exposure measure primarily captures the business-use dimension of GenAI applications rather than general consumer or leisure use. To further address the concern that we cannot fully distinguish between business and personal use, we leverage the fact that Anthropic also provides information on the proportion of queries for each task that are processed via the API. These queries are more likely to involve advanced users and may be more relevant to capture business use. Figure 13 shows that the estimated coefficient is indeed slightly higher.

Finally, estimates are robust, and the estimated effects are larger when we start the sample in 2021, thereby giving less weight to the COVID-19 pandemic period.

7.3 Alternative Mechanisms

The interpretation of our results is conditional on the assumption that our various measures of exposure to GenAI do not capture other industry and startup characteristics, which could lead to higher entry and changes in startups' size and occupational composition.

Given the robustness of our findings across alternative measures and samples, we are confident that we are capturing actual GenAI exposure. Yet, since the diffusion of GenAI technologies largely coincides with the recovery from the COVID-19 pandemic, one may wonder whether GenAI-exposed industries were severely disrupted during the COVID-19 recession. In this case, the post-2023 rise in entry could partly reflect these industries' recovery rather than the effect of GenAI.

Since, as shown in Table A.3, the top exposed industries are software, computer systems, and other IT industries, such an explanation is unlikely. Furthermore, we find that our results are, if anything, stronger when we exclude 2020 from the estimation sample. Nevertheless, to address this concern, we use industry-level exit data from the U.S. Census Bureau's Business Dynamics Statistics (BDS) to compute each four-digit NAICS industry's average exit rate in 2020–2021 and classify industries into two groups based on the severity of their pandemic-era exits. We then allow these exit-rate groups to follow their own time-varying trends. The estimated GenAI effects remain unchanged (see the last bar in Figure 12 and Figure 13), indicating that our results are not driven by differential post-COVID rebounds across industries.

8 Conclusion

We provide evidence that GenAI is spurring entrepreneurial activity, driven both by AI-native ventures building new products and services and by AI-enabled firms that lower their entry costs by automating internal operations. Although individual entrants are smaller, the rise in entry more than offsets their smaller scale: aggregate employment and earnings increase among young firms in exposed industries. Furthermore, startups appear to have higher demand for individuals with prior experience in occupations with high exposure to GenAI, which are being shed at mature companies. New ventures thus emerge as a margin through which GenAI may expand labor demand. These results alleviate concerns about widespread job displacement but also underscore the importance of prior experience. Thus, training programs for new graduates who are negatively affected by the new technology may be essential to mitigate its negative effects.

Importantly, we also document how new entrants adopt the new technology and that they overwhelmingly rely on foundation models supplied by a small number of incumbents rather than on models developed in-house, so that the core capabilities of most AI-native entrants depend on continued access to third-party APIs. Policymakers aiming to leverage the new technology to spur growth and employment should focus on ensuring reliable access to foundational models. Reducing compliance burdens and legal uncertainties for small businesses relying on GenAI, for instance, through regulatory sandboxes, may further enable economies to reap the benefits of the new technology.

References

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022). Artificial intelligence and jobs: Evidence from online vacancies. *Journal of Labor Economics 40* (S1), S293–S340.

Acharya, V. V., A. Bhardwaj, and T. Tomunen (2023). Do firms mitigate climate impact on employment? evidence from us heat shocks. Technical report, National Bureau of Economic Research.

Aghion, P. and P. Howitt (1992, March). A model of growth through creative destruction. *Econometrica 60* (2), 323–351.

Agrawal, A. K., J. McHale, and A. Oettl (2026). Enhancing worker productivity without automating tasks: A different approach to ai and the task-based model. Technical report, National Bureau of Economic Research.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135* (2), 645–709.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118* (4), 1279–1333.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics 151*, 103745.

Bao, J., B. Lou, and H. Sun (2025). Generative AI and entrepreneurial entry. *Available at SSRN 5167018*.

Bennett, V. M. and D. T. Robinson (2024). Why aren't there more minority entrepreneurs? Technical report, National Bureau of Economic Research.

Bick, A., A. Blandin, and D. Deming (2025). The state of generative ai adoption in 2025. Technical report, Federal Reserve Bank of St. Louis.

Borisov, A., A. Ellul, and M. Sevilir (2021). Access to public capital markets and employment growth. *Journal of Financial Economics 141*(3), 896–918.

Bresnahan, T. F. and M. Trajtenberg (1995). General purpose technologies 'engines of growth'? *Journal of Econometrics 65*(1), 83–108.

Brynjolfsson, E., B. Chandar, and R. Chen (2025). Canaries in the coal mine?: Six facts about the recent employment effects of artificial intelligence.

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

Brynjolfsson, E., T. Mitchell, and D. Rock (2018). What can machines learn and what does it mean for occupations and the economy? *AEA papers and proceedings 108*, 43–47.

Cai, J. J., X. Gu, L. Sheng, M. Xia, L. Zhao, and W. Zhu (2025). AI as "co-founder": GenAI for entrepreneurship. SSRN Working Paper No. 5873823.

Chen, J. and M. Ewens (2025). Venture capital and startup agglomeration. *The Journal of Finance*.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

Decker, R. A., J. Haltiwanger, R. S. Jarmin, and J. Miranda (2016). Declining business dynamism: What we know and the way forward. *American Economic Review 106*(5), 203–207. Papers and Proceedings.

Denes, M., S. T. Howell, F. Mezzanotti, X. Wang, and T. Xu (2023). Investor tax credits and entrepreneurship: Evidence from US states. *The Journal of Finance 78*(5), 2621–2671.

Dinlersoz, E., T. Dunne, J. Haltiwanger, and V. Penciakova (2021). Business formation: A tale of two recessions. *AEA Papers and Proceedings 111*, 253–257.

Eeckhout, J. and L. Veldkamp (2022). Data and market power.

Eisfeldt, A. L., G. Schubert, M. B. Zhang, and B. Taska (2024). The labor impact of generative ai on firm values. *Available at SSRN 4436627*.

Ellison, G. and E. L. Glaeser (1997). Geographic concentration in us manufacturing industries: a dartboard approach. *Journal of Political Economy 105*(5), 889–927.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2024). Gpts are gpts: Labor market impact potential of llms. *Science 384*(6702), 1306–1308.

Ewens, M. and J. Farre-Mensa (2022). Private or public equity? the evolving entrepreneurial finance landscape. *Annual Review of Financial Economics 14*(1), 271–293.

Ewens, M., R. Nanda, and M. Rhodes-Kropf (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics 128*(3), 422–442.

Felten, E., M. Raj, and R. C. Seamans (2019). The effect of artificial intelligence on human labor: An ability-based approach. *2019*(1), 15784.

Felten, E. W., M. Raj, and R. Seamans (2023). Occupational heterogeneity in exposure to generative ai. *Available at SSRN 4414065*.

Gompers, P., J. Lerner, and D. Scharfstein (2005). Entrepreneurial spawning: Public corporations and the genesis of new ventures, 1986 to 1999. *The Journal of Finance 60*(2), 577–614.

Gupta, A., F. Qian, E. Simintzi, and Y. Sun (2026). Generative AI and entrepreneurship. Working Paper, UNC Kenan-Flagler Business School.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013). Who creates jobs? small versus large versus young. *The Review of Economics and Statistics 95*(2), 347–361.

Hampole, M., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2025). Artificial intelligence and the labor market. Technical report, National Bureau of Economic Research.

Humlum, A. and E. Vestergaard (2024). The adoption of ChatGpt.

Humlum, A. and E. Vestergaard (2025). Large language models, small labor market effects. Technical report, National Bureau of Economic Research.

Jiang, W., J. Park, R. J. Xiao, and S. Zhang (2025). AI and the extended workday: Productivity, contracting efficiency, and distribution of rents. Technical report, National Bureau of Economic Research.

Jones, C. I. (2024). The AI dilemma: Growth versus existential risk. *American Economic Review: Insights 6*(4), 575–90.

Jovanovic, B. and P. L. Rousseau (2005). General purpose technologies. In *Handbook of Economic Growth*, Volume 1, pp. 1181–1224. Elsevier.

Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of political economy 112*(5), 986–1018.

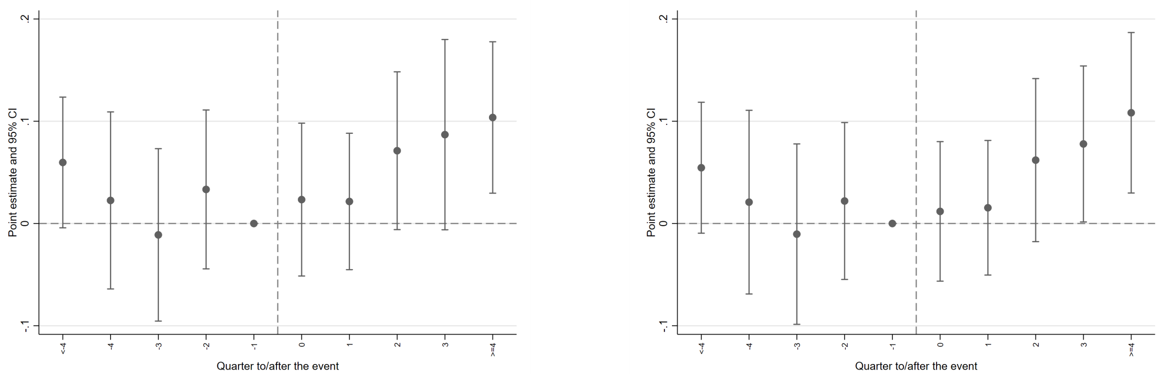
Lu, Y., G. M. Phillips, and J. Yang (2024). The impact of cloud computing and ai on industry dynamics and concentration. Technical report, National Bureau of Economic Research.

Marchesi, E. and S. Tang (2026). When AI meets entrepreneurship: Evidence from the commercialization of ChatGPT. Working Paper.

Nanda, R. and J. B. Sørensen (2010). Workplace peers and entrepreneurship. *Management science 56*(7), 1116–1126.

Shao, Y., H. Zope, Y. Jiang, J. Pei, D. Nguyen, E. Brynjolfsson, and D. Yang (2025). Future of work with ai agents: Auditing automation and augmentation potential across the us workforce. *arXiv preprint arXiv:2506.06576*.

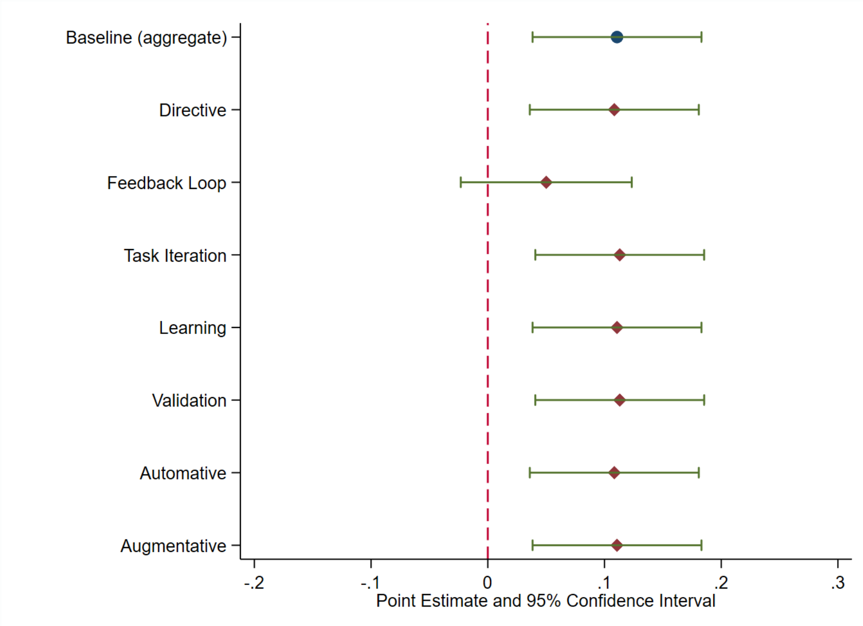
Figure 1: Dynamic Effects of GenAI Introduction on Business Formation



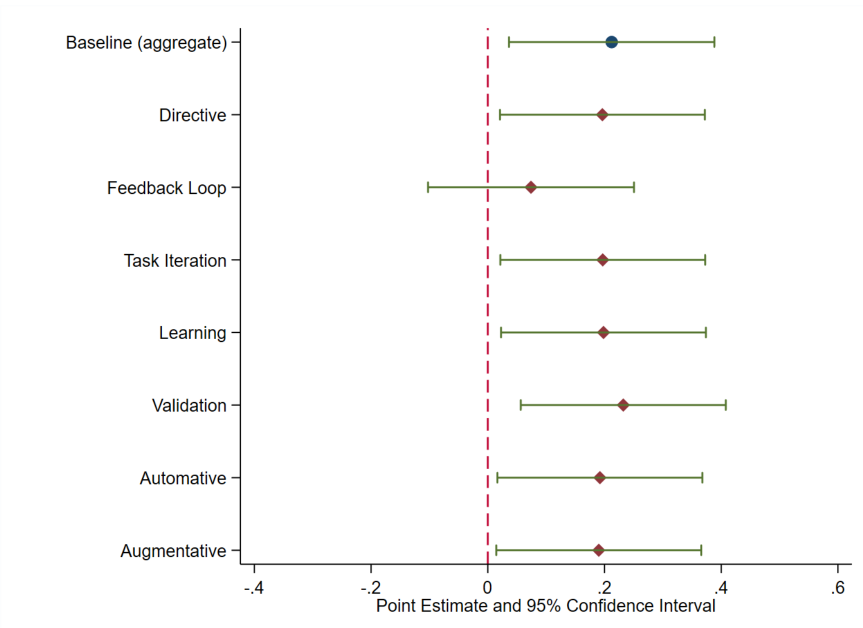
(a) Within 4 Quarters (PBF4Q) **(b) Within 8 Quarters** (PBF8Q)

This figure illustrates the dynamic effects of GenAI introduction on projected business formations within four quarters (Panel a) and eight quarters (Panel b). To reduce estimation noise, months are aggregated into quarterly intervals. Quarter -1 corresponds to the second quarter of 2022 and serves as the omitted baseline period. Panel a plots the event-time coefficients from the dynamic specification corresponding to Column (4) in Table 2, and Panel b presents the analogous results for Column (5).

Figure 2: Effects of GenAI Introduction on Firm Entry—Types of Interaction



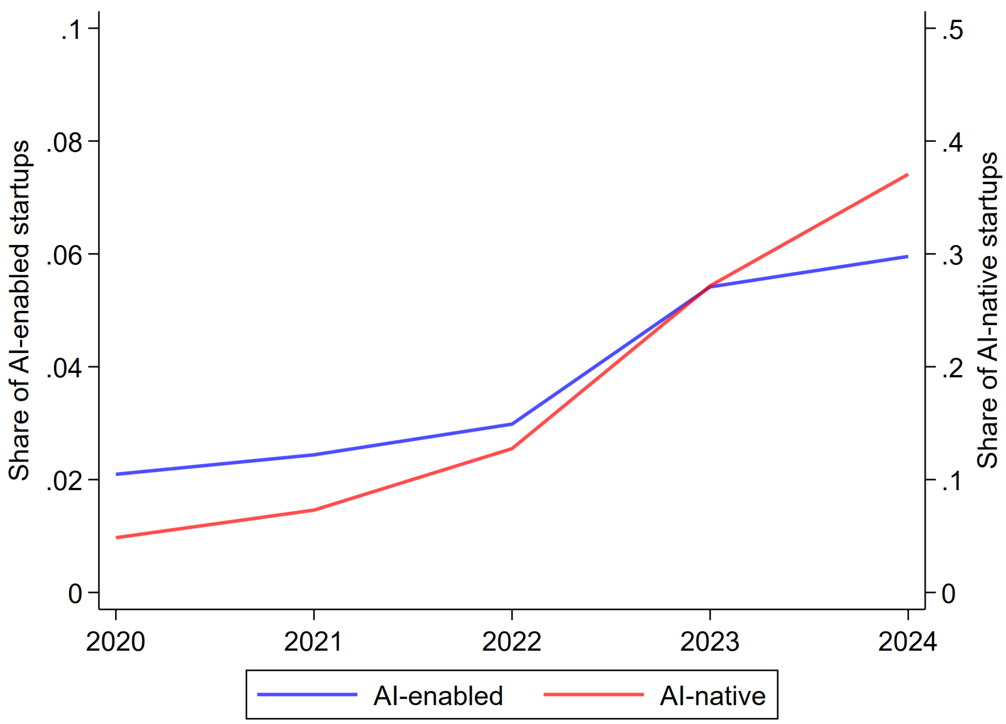
(a) Probability of Firm Entry



(b) # Firm Entry (log)

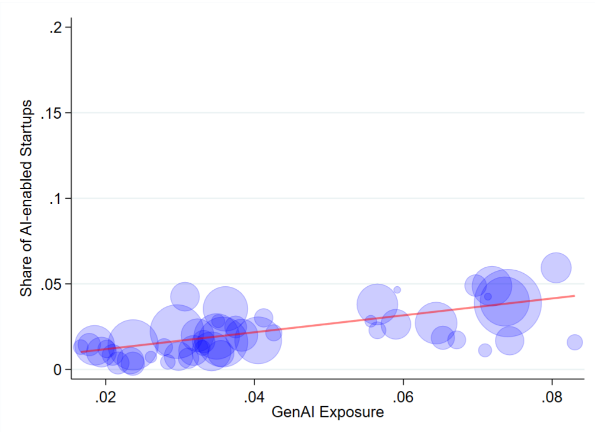
This figure decomposes the effect of GenAI introduction on firm entry by type of user–AI interaction, using the Claude.ai conversation data. The analysis distinguishes among five interaction types: *directive*, *feedback loop*, *task iteration*, *learning*, *validation*. These categories capture distinct modes of GenAI use, ranging from automation-oriented interactions (directive, feedback loop) to augmentative and exploratory ones (validation, task iteration, learning). Panel (a) reports the estimated effect on the probability of firm entry, and Panel (b) reports the effect on the (log) number of new firm entries. The specification in Panel (a) is analogous to Column (2) in Table 3, and the specification in Panel (b) corresponds to Column (4) in Table 3.

Figure 3: Share of AI-Enabled and AI-Native Startups

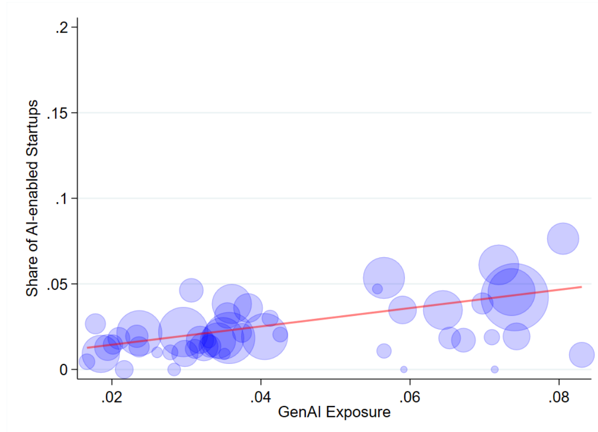


This figure plots the share of startups classified as AI-enabled and AI-native each year, from 2020 to 2024, using deep research.

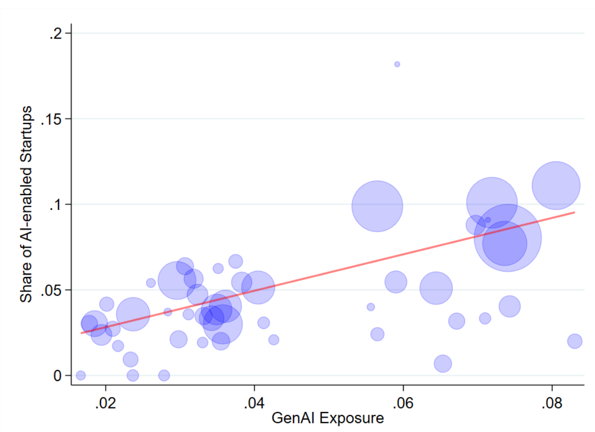
Figure 4: AI-Enabled Startups and Exposure to GenAI



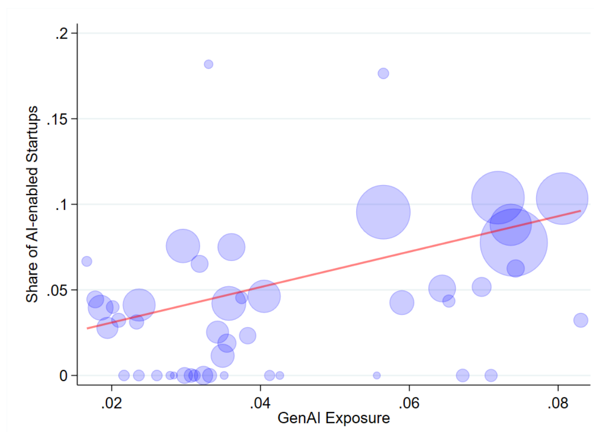
(a) 2020



(b) 2021



(c) 2023

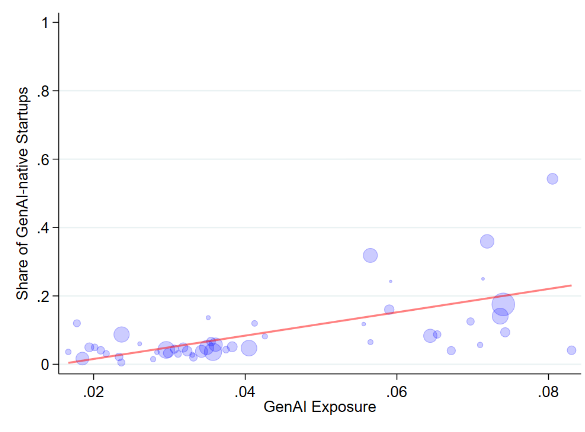


(d) 2024

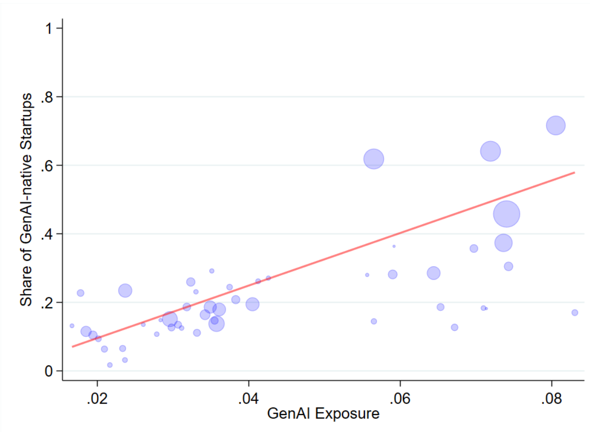
This figure plots the share of AI-enabled startups (identified using deep research) in each category group against the group's exposure to GenAI for the years before (2020-2021) and after (2023-2024) the introduction of GenAI. Each point represents a category group, with bubble size proportional to the total number of startups in that group in the corresponding year.

Figure 5: AI-Native Startups and Exposure to GenAI

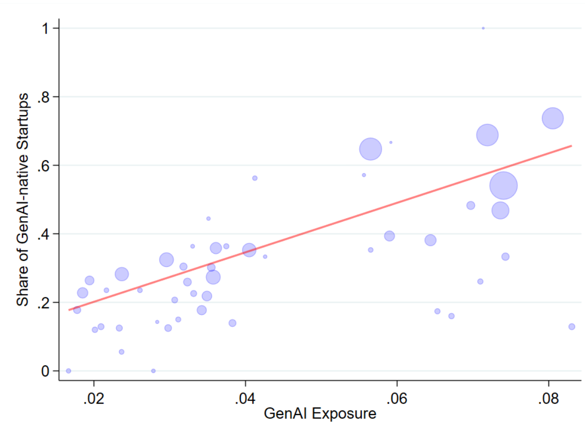
(a) 2020



(b) 2021



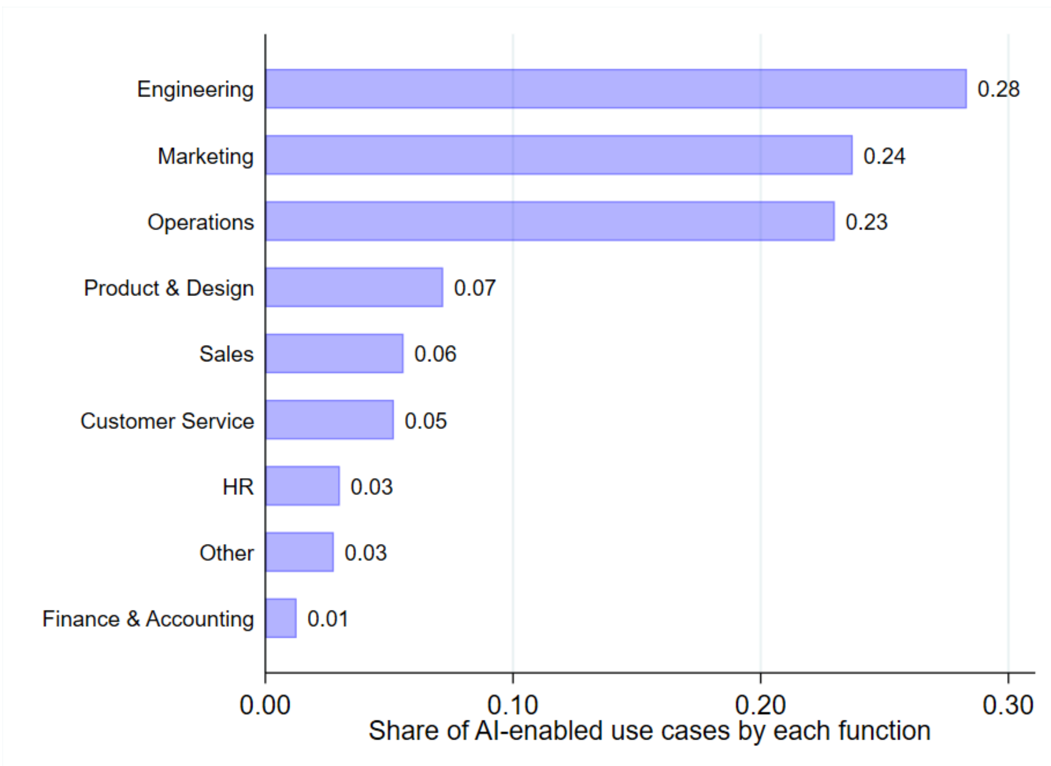
(c) 2023



(d) 2024

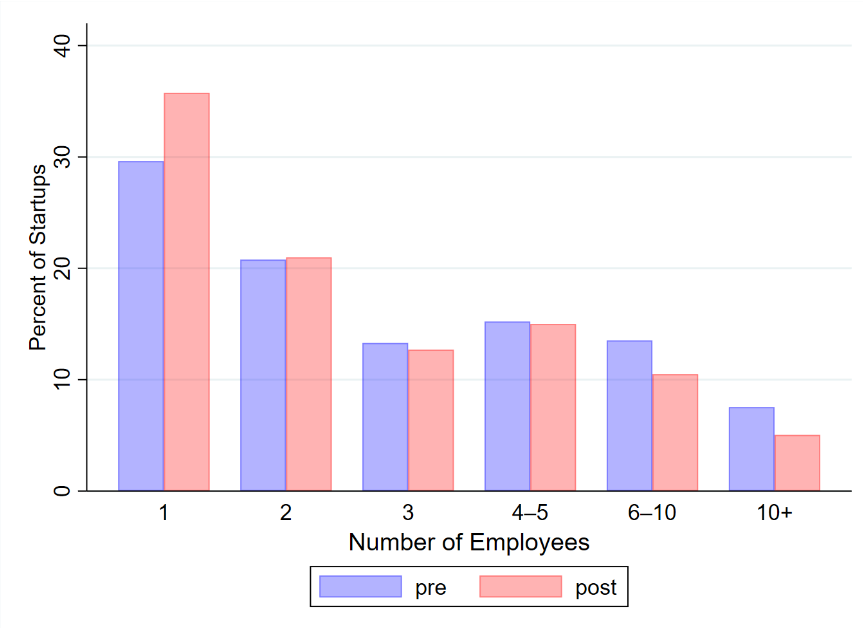
This figure plots the share of AI-native startups (identified using deep research) in each category group against the group's exposure to GenAI for the years before (2020–2021) and after (2023–2024) the introduction of GenAI. Each point represents a category group, with bubble size proportional to the total number of startups in that group in the corresponding year.

Figure 6: Most Frequent AI-Enabled Business Functions

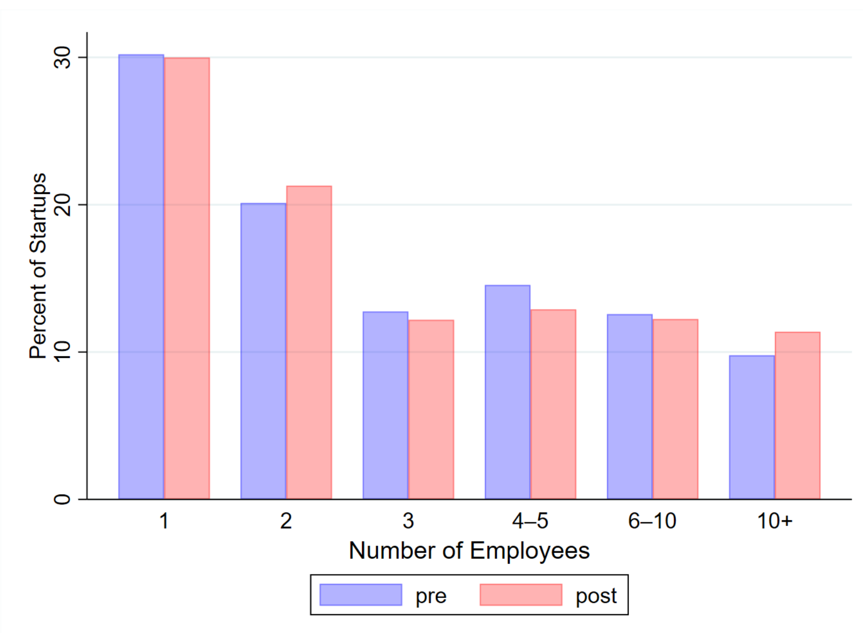


This figure displays the most common AI-enabled functions identified among AI-enabled startups.

Figure 7: GenAI Introduction and the Changing Team Size Distribution of Startups



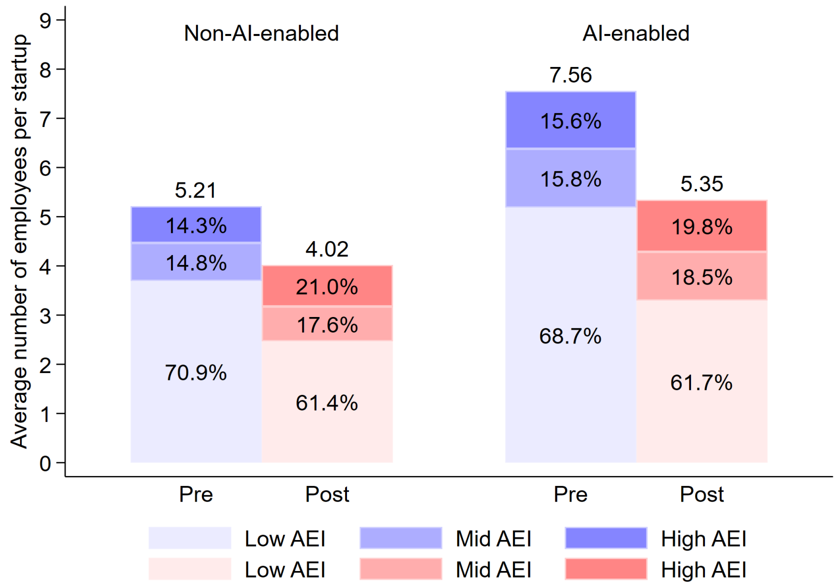
(a) High GenAI Exposure



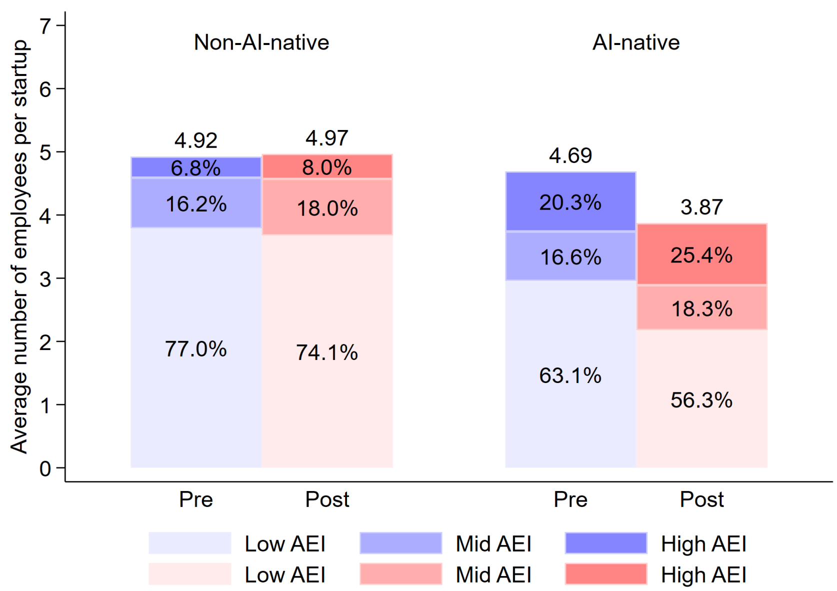
(b) Low GenAI Exposure

This figure shows the distribution of startups by team size—measured as the number of employees at the end of the founding year—before and after the introduction of GenAI. Panel (a) presents startups in industries with high GenAI exposure, while Panel (b) presents those in low-exposure industries.

Figure 8: Employee Composition of AI-Enabled and AI-Native Startups



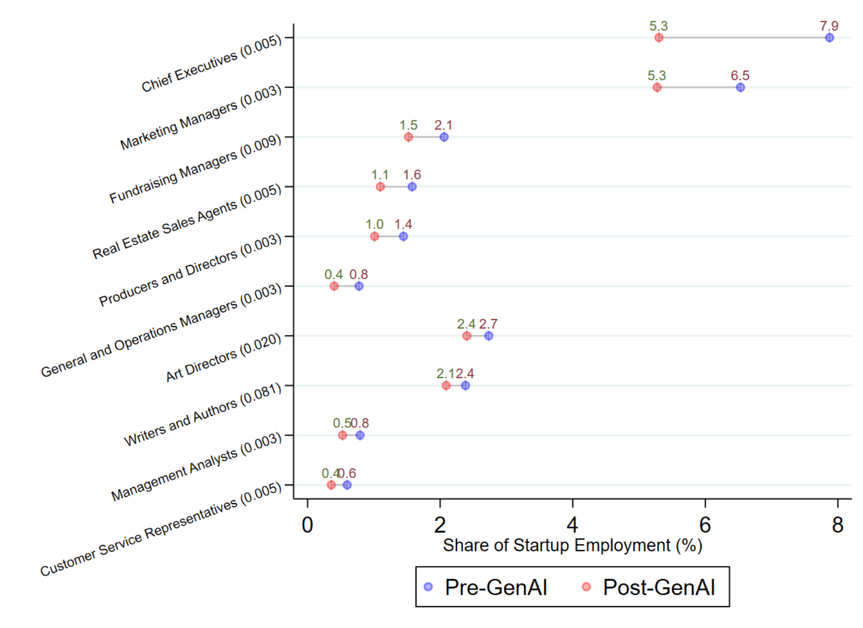
(a) AI-Enabled Startups



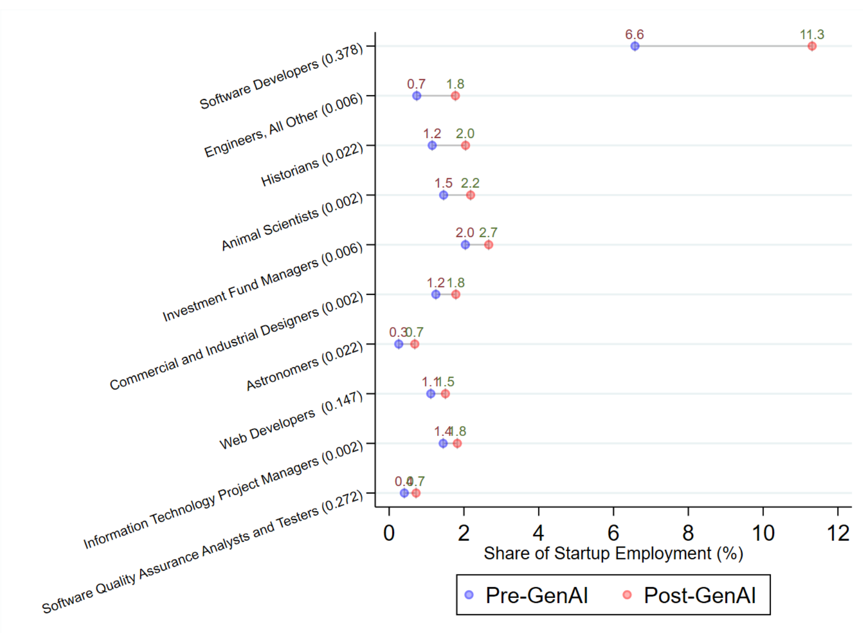
(b) AI-Native Startups

This figure plots the average number of employees per startup at the end of the founding year—before and after the introduction of GenAI, separately by AI-enabled and AI-native classification. Panel (a) compares AI-enabled startups with non-AI-enabled startups. Panel (b) compares AI-native startups with non-AI-native startups. Each stacked bar reports the average number of employees per startup, decomposed into low-, mid-, and high-AEI employee groups. The numbers above the bars report the total average number of employees per startup. The percentages inside each segment report the share of each AEI group in total employment within the corresponding startup group and period.

Figure 9: GenAI Introduction and Changing Startup Occupations



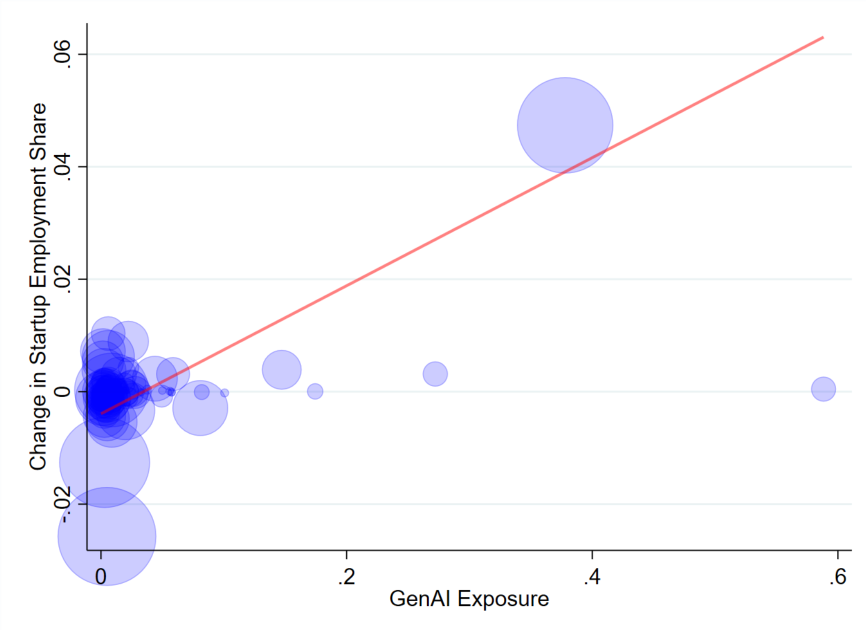
(a) Losing Occupations



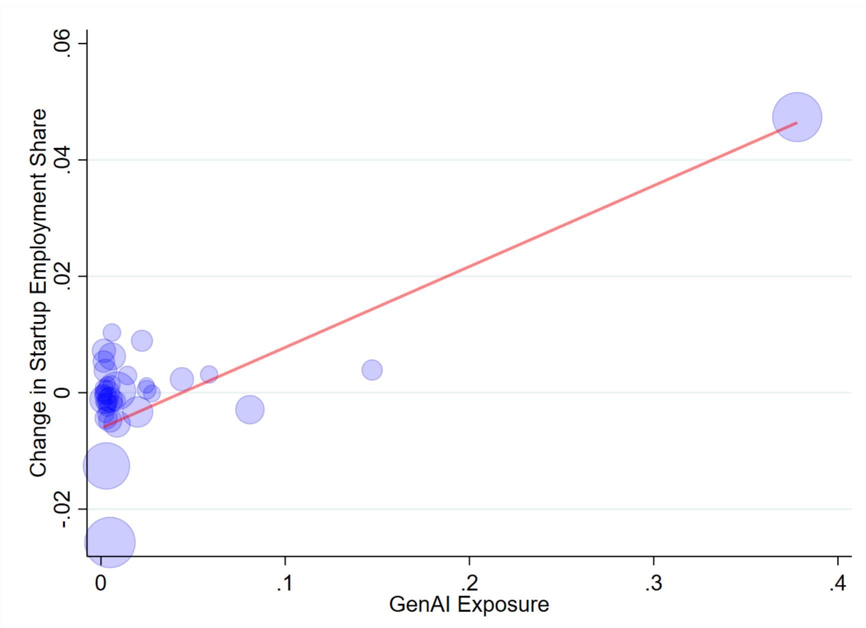
(b) Gaining Occupations

This figure shows the occupations that experienced the largest changes in their employment shares among startups following the introduction of GenAI. Panel (a) lists the ten occupations with the largest declines in employment share, while Panel (b) lists those with the largest gains.

Figure 10: Changes in Startup Occupations vs. GenAI Exposure



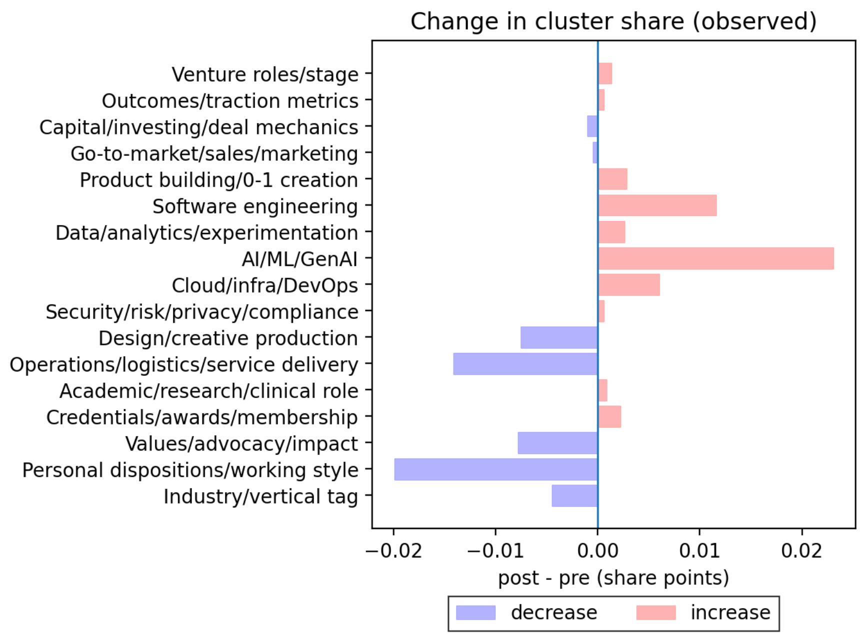
(a) All Occupations



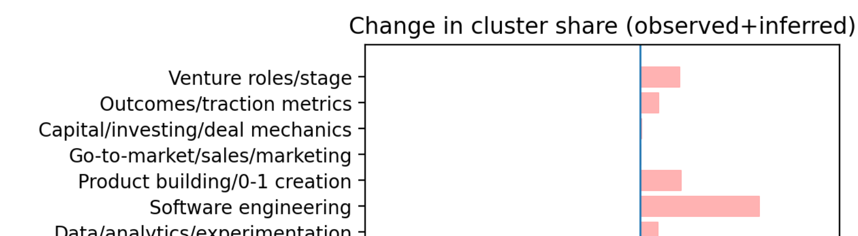
(b) Important Occupations (top decile)

This figure plots the change in startup employment share at the occupation level before and after the introduction of GenAI against each occupation's GenAI exposure measure. Panel (a) includes all occupations, while Panel (b) focuses on the top decile of occupations ranked by total employment share. Each point represents an occupation, with bubble size proportional to total employment, and the fitted line illustrates the relationship between exposure and employment share changes.

Figure 11: GenAI Introduction and Founders' Changing Traits



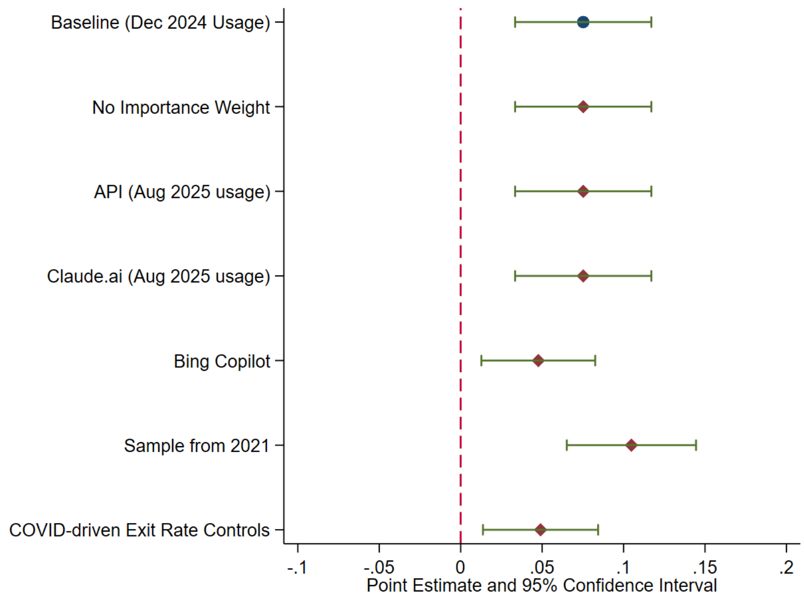
(a) Observed Traits



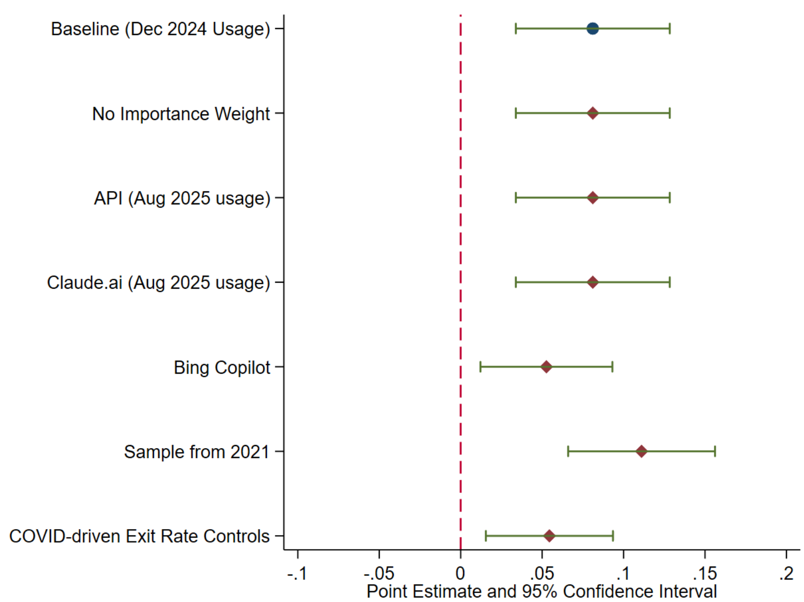
(b) Observed + Inferred Traits

This figure shows changes in the distribution of founder traits following the diffusion of GenAI, based on traits extracted from founders' unstructured self-descriptions at the time of entry. Traits are organized into bottom-up clusters that capture both technical, task-based capabilities and broader dispositional or values-oriented attributes (see Appendix C for details). Panel (a) reports results using traits explicitly stated in founders' profiles, while Panel (b) additionally incorporates traits logically inferred from founders' narratives and actions.

Figure 12: Effects of GenAI Introduction on Business Formation—Robustness



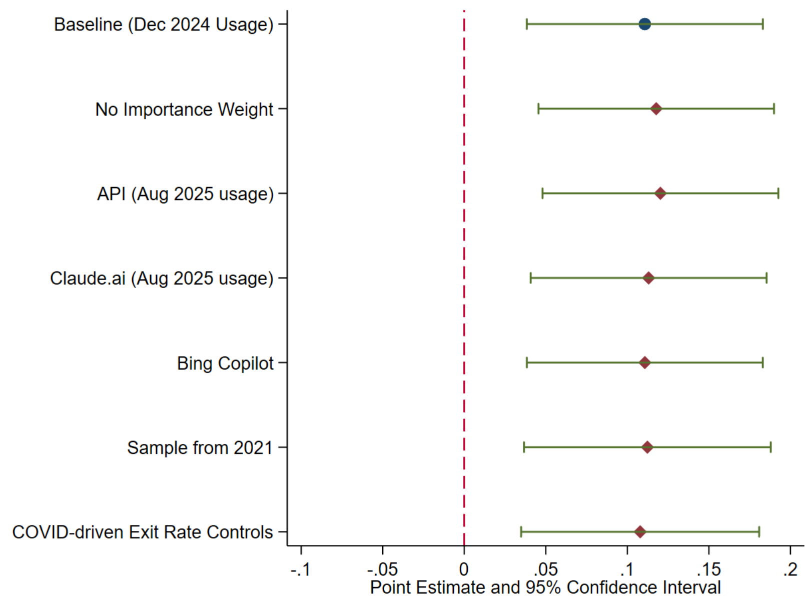
(a) **Within 4 Quarters** (PBF4Q)



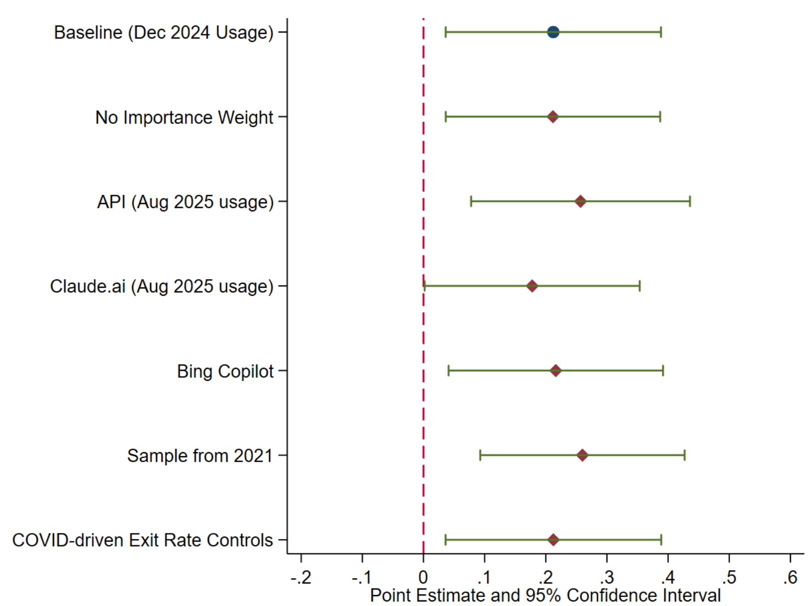
(b) **Within 8 Quarters** (PBF8Q)

This figure presents robustness tests for the baseline estimates of the effect of GenAI introduction on projected business formations within four quarters (Panel a) and eight quarters (Panel b). The panels replicate the baseline specification using alternative measures of industry exposure to generative AI and alternative sample periods. The specification in Panel (a) is analogous to Column (4) in Table 2, and the specification in Panel (b) corresponds to Column (5) in Table 2.

Figure 13: Effects of GenAI Introduction on Firm Entry—Robustness



(a) Probability of Firm Entry



(b) # Firm Entry (log)

This figure presents robustness tests for the baseline estimates of the effect of GenAI introduction on firm entry. The panels replicate the baseline specification using alternative measures of industry exposure to generative AI and alternative sample periods. Panel (a) reports the estimated effect on the probability of firm entry, and Panel (b) reports the effect on the (log) number of new firm entries. The specification in Panel (a) is analogous to Column (2) in Table 3, and the specification in Panel (b) corresponds to Column (4) in Table 3.

Table 1: Summary Statistics

Panel A: Sector-Year-Month Level (BFS) and Industry-Year-Quarter Level (BDS)

Variable	N	Mean	Std. Dev.	Median	p1	p25	p75	p99
# BA	1,105	23,961	20,311	21,052	362	8,162	34,412	85,976
# CBA	1,105	2,838	2,704	2,335	30	1,030	3,837	8,575
# HBA	1,105	8,076	7,969	5,190	54	2,051	15,215	26,770
# PBF4Q	1,099	1,528	1,269	1,159	15	476	2,668	4,142
# PBF8Q	1,099	2,109	1,746	1,547	21	669	3,749	5,658
AEI	1,105	0.009	0.011	0.005	0.001	0.003	0.008	0.039

Panel B: Industry-Year Level (aggregated from InfoGroup)

Variable	N	Mean	Std. Dev.	Median	p1	p25	p75	p99
I (firm entry)	1,430	0.778	0.416	1.000	0.000	1.000	1.000	1.000
# firm entry	1,430	63.096	169.319	7.000	0.000	1.000	45.000	741.000
# firm entry (below median size)	1,430	34.311	98.802	3.000	0.000	0.000	23.000	453.000
# firm entry (above median size)	1,430	28.785	82.546	2.000	0.000	0.000	18.000	368.000
# state	1,430	12.779	14.972	5.000	0.000	1.000	22.000	49.000
# csa	1,430	17.481	25.544	5.000	0.000	1.000	24.000	103.000
# county	1,430	41.479	84.113	6.000	0.000	1.000	41.000	394.000
HHI state	1,430	0.441	0.408	0.222	0.042	0.080	1.000	1.000
HHI csa	1,430	0.458	0.403	0.250	0.053	0.093	1.000	1.000
HHI county	1,430	0.408	0.429	0.167	0.006	0.028	1.000	1.000
AEI	1,430	0.009	0.015	0.004	0.001	0.002	0.008	0.073

Panel C: Startup Level (Crunchbase-LinkedIn Merged Sample)

Variable	N	Mean	Std. Dev.	Median	p1	p25	p75	p99
# employee	42,537	4.592	7.325	2.000	1.000	1.000	5.000	39.000
# occupation	42,537	3.345	3.839	2.000	1.000	1.000	4.000	20.000
HHI occupation	42,537	0.570	0.333	0.500	0.074	0.280	1.000	1.000
# employee (high AEI)	42,537	0.434	1.333	0.000	0.000	0.000	0.000	5.000
# employee (mid AEI)	42,537	0.769	1.648	0.000	0.000	0.000	1.000	7.000
# employee (low AEI)	42,537	3.389	6.045	2.000	0.000	1.000	4.000	31.000
AEI	42,537	0.044	0.016	0.039	0.019	0.031	0.058	0.074

Panel D: Founder Level (Crunchbase-LinkedIn Merged Sample)

Variable	N	Mean	Std. Dev.	Median	p1	p25	p75	p99
Prior AEI (most recent)	15,236	0.046	0.111	0.005	0.000	0.003	0.019	0.378
Prior AEI (mean)	15,236	0.047	0.094	0.008	0.000	0.004	0.023	0.378
Prior AEI (max)	15,236	0.092	0.157	0.020	0.000	0.006	0.062	0.588
Startup AEI	15,236	0.055	0.088	0.011	0.001	0.005	0.070	0.378
Previous-Job $\geq$Manager	14,615	0.708	0.455	1.000	0.000	0.000	1.000	1.000
Previous-Job Seniority	14,615	4.453	1.770	5.000	1.000	3.000	6.000	7.000
Experience	15,236	15.065	8.611	13.833	0.000	8.583	20.417	60.667
Log Previous-Job Compensation	14,615	12.313	0.921	12.364	7.779	11.665	12.940	15.023
Paycut (from Previous-Job, 0/1)	12,629	0.212	0.409	0.000	0.000	0.000	0.000	1.000
Education	13,757	3.924	1.177	4.000	1.000	3.000	5.000	6.000
Male	17,402	0.781	0.370	0.995	0.000	0.719	0.997	1.000

This table shows summary statistics for different samples. Panel A summarizes sector-year-month level counts of new business registrations from BFS. Panel B summarizes industry–year–level variables constructed from the InfoGroup firm panel, aggregated to the four-digit NAICS level. Panel C presents summary statistics for startup-level variables from the Crunchbase–LinkedIn merged sample. Panel D reports founder-level characteristics from the same Crunchbase–LinkedIn sample.

Table 2: Effects of GenAI Introduction on Business Formation

	(1)	(2)	(3)	(4)	(5)
Dep. Var.	# new business formation (log)				
	BA	HBA	CBA	PBF4Q	PBF8Q
AEI (above median) × Post_2022m8	0.038**	0.038**	0.055**	0.056**	0.062**
	[0.016]	[0.017]	[0.022]	[0.022]	[0.025]
Sector FE	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES
Obs	1,105	1,105	1,105	1,099	1,099
Adj R2	0.992	0.993	0.987	0.987	0.983

This table reports the estimated effects of GenAI introduction on new business formation using aggregate-level data from the U.S. Census Bureau's *Business Formation Statistics* (BFS). The outcome variable is the logarithm of the number of new business formations in a two-digit NAICS industry, measured monthly. The sample covers January 2020 to April 2025. AEI (above median) is a dummy that equals one for sectors whose AI exposure, measured by the Anthropic Economic Index (AEI) aggregated to the two-digit NAICS level, is above median, and zero otherwise. Post_2022m8 is a dummy that equals 1 after August, 2022, and 0 otherwise. Columns (1)-(5) use five BFS outcome measures: (1) all Business Applications (BA), (2) High-Propensity Business Applications (HBA), (3) Business Applications from Corporations (CBA), (4) Projected Business Formations within four quarters (PBF4Q), and (5) Projected Business Formations within eight quarters (PBF8Q). HBA represent EIN applications with characteristics historically associated with the successful establishment of active firms, such as intended payroll or corporate registration. PBF measures are Census-derived projections of business applications expected to become active employer firms within four to eight quarters and are considered a leading indicator of realized business creation. Industry (two-digit NAICS) and month-year fixed effects are included in all specifications. Standard errors are obtained via a bootstrap procedure with 500 repetitions, resampling at the sector level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 3: Effects of GenAI Introduction on Firm Entry

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
AEI (quartile) × Post_2022	0.040**		0.055*	
	[0.017]		[0.033]	
AEI (above median) × Post_2022		0.111***		0.212**
		[0.037]		[0.089]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.471	0.473	0.888	0.888

This table reports the estimated effects of GenAI introduction on new firm entry using establishment-level data aggregated from the *InfoGroup* firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. The dependent variable in Columns (1) and (2) is an indicator for any firm entry, or *I(firm entry)*. The dependent variable in Columns (3) and (4) is the logarithm of one plus the number of firm entries in each industry–year cell, or *# firm entry (log)*. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (quartile)* denotes the four-digit NAICS-level AI exposure measure based on AEI, divided into quartiles according to the exposure distribution across industries. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 4: Effects of GenAI Introduction on the Geographic Distribution of Firm Entry

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	# state	# csa	# county	HHI state	HHI csa	HHI county
AEI (above median) × Post_2022	0.204***	0.197***	0.221***	-0.064**	-0.058**	-0.068***
	[0.069]	[0.072]	[0.084]	[0.025]	[0.024]	[0.026]
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430	1,430	1,430
Adj R2	0.854	0.868	0.884	0.764	0.762	0.776

This table reports the estimated effects of GenAI introduction on the geographic distribution of new firm entries, using establishment-level data aggregated from the InfoGroup firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. Columns 1-3 examine the extensive margin of geographic dispersion. The dependent variables are measured as the logarithm of one plus the number of distinct geographic units—states, combined statistical areas (CSAs), and counties—recording new firm entries within each industry–year cell. Columns 4-6 examine geographic concentration, measured by the Herfindahl–Hirschman Index (HHI) of firm entry across states, CSAs, and counties. Lower HHI values indicate more geographically dispersed firm creation. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 5: GenAI Introduction and the Geographic Dispersion of Entry—Channels

Panel A: Relaxed Labor Specialization Requirements

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
	non-specialized	specialized	non-specialized	specialized
AEI (above median) × Post_2022	0.110***	-0.028	0.226**	-0.155***
	[0.037]	[0.037]	[0.089]	[0.059]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.48	0.571	0.886	0.85

Panel B: Relaxed Financing Requirements

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
	non VC hubs	VC hubs	non VC hubs	VC hubs
AEI (above median) × Post_2022	0.101***	0.057	0.224***	-0.006
	[0.038]	[0.037]	[0.083]	[0.077]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.489	0.587	0.89	0.86

This table examines two channels through which the introduction of GenAI may have contributed to more geographically dispersed firm entry, using establishment-level data aggregated from the InfoGroup firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. Panel A investigates the labor specialization channel. Each county is classified as having *specialized* or *non-specialized* labor in a given industry based on the pre-GenAI distribution of local employment. For each industry, the share of workers employed in that industry relative to total county employment is calculated, and counties in the top decile of this distribution are defined as specialized, while all others are classified as non-specialized. Panel B tests the venture capital financing channel. We classify regions into *VC hubs* and *non-hubs* based on pre-GenAI venture investment intensity. VC hubs correspond to the three states with the highest levels of venture capital activity—California, Massachusetts, and New York—while all other states are treated as non-hubs. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 6: Effects of GenAI Introduction on the Size Composition of Firm Entry

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
	below median size	above median size	below median size	above median size
AEI (above median) × Post_2022	0.108***	0.046	0.231***	-0.094
	[0.040]	[0.032]	[0.082]	[0.072]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.468	0.628	0.857	0.916

This table reports the estimated effects of GenAI introduction on the size distribution of new firms, using establishment-level data aggregated from the *InfoGroup* firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. The dependent variable is the logarithm of one plus the number of new firm entries (*# firm entry (log)*) in each industry–year cell, separately for firms with initial employment below or above the pre-treatment (2020-2022) industry median employment. Columns (1) and (3) examine firms below the median size threshold, and Columns (2) and (4) examine those above it. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 7: Effects of GenAI Introduction on Startup Team Size and Occupations

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	# employee	# occupation	HHI occupation	# employee	# occupation	HHI occupation
AEI (above median) × Post_2022	-1.107***	-0.560***	0.051***	-0.978***	-0.485***	0.045***
	[0.257]	[0.118]	[0.010]	[0.249]	[0.129]	[0.010]
Post_2022	0.423*	0.145	-0.001			
	[0.232]	[0.106]	[0.008]			
AEI (above median)	-0.531***	-0.233***	-0.003	0.05	0.126*	-0.021***
	[0.076]	[0.039]	[0.003]	[0.140]	[0.072]	[0.006]
Industry Category Indicators	NO	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES
Obs	42,537	42,537	42,537	42,537	42,537	42,537
Adj R2	0.003	0.003	0.002	0.011	0.018	0.018

This table reports the estimated effects of GenAI introduction on startup team size and occupational composition, using firm-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is a startup firm founded between 2020 and 2024, with all dependent variables measured at the end of the founding year. Columns (1) and (4) report the total number of employees; Columns (2) and (5) report the number of unique occupations within the firm; and Columns (3) and (6) report the Herfindahl–Hirschman Index (HHI) of occupational concentration, which captures the degree of within-firm specialization. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for firms operating in category groups with above-median GenAI exposure, and zero otherwise. Columns (1)–(3) report results without category-group controls or year fixed effects, while Columns (4)–(6) include both sets of controls. Standard errors are obtained via a bootstrap procedure with 500 repetitions. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 8: Effects of GenAI Introduction on Aggregate Employment and Earnings by Firm Age

	(1)	(2)	(3)	(4)
Dep. Var.	employment (log)		earnings	
Firm Age	0−1	>1	0−1	>1
AEI (above median) × Post_2022	0.074*	0.034	222.390**	-20.9
	[0.043]	[0.023]	[106.796]	[51.887]
Dep. Var. Mean	8.153	9.352	4624.199	5248.697
Industry FE	YES	YES	YES	YES
Year-Quarter FE	YES	YES	YES	YES
Obs	5,218	20,890	5,211	20,880
Adj R2	0.973	0.553	0.829	0.801

This table reports the estimated effects of GenAI introduction on industry-level employment and earnings using the Census Quarterly Workforce Indicators (QWI). The unit of observation is a four-digit NAICS industry–quarter. The dependent variables are (i) the logarithm of employment for age 0–1 firms and for all older firms in the industry, and (ii) average monthly earnings for the same two age groups. Columns (1)–(2) present results for employment; Columns (3)–(4) present results for earnings. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 9: Effects of GenAI Introduction on Startup Employment by Occupational Exposure

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	# employee (by occupation)					
	high AEI	mid AEI	low AEI	high AEI	mid AEI	low AEI
AEI (above median) × Post_2022	0.068*	-0.285***	-0.890***	0.022	-0.241***	-0.760***
	[0.039]	[0.054]	[0.202]	[0.040]	[0.053]	[0.204]
Post_2022	0.050***	0.186***	0.188			
	[0.015]	[0.047]	[0.185]			
AEI (above median)	0.561***	0.044**	-1.136***	0.186***	0.130***	-0.266**
	[0.013]	[0.018]	[0.064]	[0.027]	[0.033]	[0.109]
Industry Category Indicators	NO	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES
Obs	42,537	42,537	42,537	42,537	42,537	42,537
Adj R2	0.047	0.001	0.012	0.064	0.037	0.025

This table reports the estimated effects of GenAI introduction on startup employment across occupations with differing exposure to GenAI-related tasks, using firm-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is a startup firm founded between 2020 and 2024, with all dependent variables measured at the end of the founding year. Occupations are grouped into three categories based on their Anthropic Economic Index (AEI) values: high exposure ($AEI > 0.1$) in Columns (1) and (4), medium exposure ($0.01 \leq AEI \leq 0.1$) in Columns (2) and (5), and low exposure ($AEI < 0.01$) in Columns (3) and (6). The dependent variable is the number of employees in each AEI category within a firm. Columns (1)–(3) report results without category-group controls or year fixed effects, while Columns (4)–(6) include both sets of controls. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for firms operating in category groups with above-median GenAI exposure, and zero otherwise. Columns (1)–(3) report results without category-group controls or year fixed effects, while Columns (4)–(6) include both sets of controls. Standard errors are obtained via a bootstrap procedure with 500 repetitions. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 10: Effects of GenAI Introduction on Startup Founders' Occupational Backgrounds

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	AEI of past positions					
	Most Recent	Mean	Max	Most Recent	Mean	Max
AEI (above median) × Post_2022	0.021***	0.013***	0.028***	0.021***	0.013***	0.027***
	[0.005]	[0.004]	[0.007]	[0.005]	[0.004]	[0.007]
Post_2022	0.000	-0.000	-0.001			
	[0.003]	[0.002]	[0.004]			
AEI (above median)	0.045***	0.047***	0.085***	0.040***	0.041***	0.073***
	[0.002]	[0.001]	[0.002]	[0.003]	[0.002]	[0.004]
Dep. Var. Mean	0.047	0.047	0.092	0.047	0.047	0.092
Industry Category Indicators	NO	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES
Obs	14,787	14,787	14,787	14,787	14,787	14,787
Adj R2	0.048	0.065	0.081	0.050	0.068	0.085

This table examines how the occupational backgrounds of startup founders differ across industries with varying exposure to GenAI, using founder-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is an individual founder of startups founded between 2020 and 2024. The dependent variables measure how exposed each founder's pre-founding career was to GenAI-related technological change. Specifically, Columns (1) and (4) use the AEI of the founder's most recent occupation prior to founding, Columns (2) and (5) use the mean AEI across all pre-founding occupations, and Columns (3) and (6) use the maximum AEI observed across the founder's career history. Higher values indicate that the founder's prior jobs were more susceptible to GenAI-related automation or transformation. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for firms operating in category groups with above-median GenAI exposure, and zero otherwise. Columns (1)–(3) report estimates without category-group controls or year fixed effects, while Columns (4)–(6) include both sets of controls. Standard errors are obtained via a bootstrap procedure with 500 repetitions. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 11: Effects of GenAI Introduction on Startup Founder Traits

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Var.	Capital/ Investing	AI/ML GenAI	Product Manager/ Development	Marketing/ Sales	Operations/ Logistics	Capital/ Investing	AI/ML GenAI	Product Manager/ Development	Marketing/ Sales	Operations/ Logistics
AEI (above median) × Post_2022	-0.134**	0.247***	-0.079**	-0.060	-0.070	-0.131**	0.215***	-0.084**	-0.048	-0.063
	[0.056]	[0.037]	[0.037]	[0.083]	[0.073]	[0.056]	[0.037]	[0.037]	[0.083]	[0.073]
Post_2022	0.111**	0.036	0.056*	0.064	0.020					
	[0.048]	[0.024]	[0.030]	[0.070]	[0.062]					
AEI (above median)	-0.142***	0.210***	0.249***	0.031	-0.191***	-0.162***	0.013	0.238***	0.106***	-0.117***
	[0.017]	[0.009]	[0.012]	[0.023]	[0.023]	[0.021]	[0.012]	[0.017]	[0.032]	[0.029]
Total observed traits	0.005***	0.001***	0.005***	0.013***	0.053***	0.005***	0.001**	0.005***	0.013***	0.053***
	[0.001]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]	[0.000]	[0.000]	[0.001]	[0.001]
Dep. Var. Mean	0.425	0.217	0.386	0.700	1.096	0.425	0.217	0.386	0.700	1.096
Industry Category Indicators	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES
Year FE	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES
Obs	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154
Adj R2	0.013	0.051	0.034	0.021	0.272	0.015	0.079	0.034	0.023	0.272

This table reports the estimated effects of GenAI introduction on observed founder trait counts, using founder-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is an individual founder of startups founded between 2020 and 2024. The dependent variables are counts of observed traits in five trait clusters defined in the appendix, Appendix C: Cluster 3, capital, investing, and deal mechanics; Cluster 8, AI/ML and GenAI; Cluster 5, product building and "0–1" creation; Cluster 4, go-to-market, including marketing, sales, growth, and partnerships; and Cluster 12, operations, service delivery, and logistics. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for firms operating in category groups with above-median GenAI exposure, and zero otherwise. All columns control for the total number of observed traits extracted for the founder across the 16 trait groups. Columns (1)–(5) report estimates without category-group controls or year fixed effects, while Columns (6)–(10) include both sets of controls. Standard errors are obtained via a bootstrap procedure with 500 repetitions. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table 12: Effects of GenAI Introduction on Startup Founder Previous Employment and Characteristics

Panel A: No Controls

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	Expand	Shrink	≥Manager	Seniority	Experience	Compensation	Paycut	Education	Male
AEI (above median) × Post_2022	0.046**	0.025	-0.055**	-0.197**	0.350	-0.012	-0.021	-0.113*	-0.025
	[0.019]	[0.021]	[0.025]	[0.099]	[0.504]	[0.053]	[0.029]	[0.067]	[0.018]
Post_2022	-0.041***	-0.015	0.055***	0.248***	0.868**	0.126***	0.148***	0.108*	0.026
	[0.014]	[0.016]	[0.020]	[0.082]	[0.421]	[0.043]	[0.025]	[0.058]	[0.016]
AEI (above median)	0.008	0.038***	-0.033***	-0.091***	-1.724***	-0.060***	-0.009	-0.149***	0.080***
	[0.007]	[0.006]	[0.008]	[0.032]	[0.153]	[0.017]	[0.008]	[0.022]	[0.006]
Dep. Var. Mean	0.100	0.098	0.706	4.443	14.989	12.307	0.209	3.926	0.781
Industry Category Indicators	NO	NO	NO	NO	NO	NO	NO	NO	NO
Year FE	NO	NO	NO	NO	NO	NO	NO	NO	NO
Obs.	8,826	8,826	14,284	14,284	14,787	14,284	12,343	13,370	16,906
Adj. R^2	0.001	0.004	0.002	0.001	0.010	0.002	0.010	0.004	0.011

Panel B: Category Controls and Founded-Year Fixed Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. Var.	Expand	Shrink	≥Manager	Seniority	Experience	Compensation	Paycut	Education	Male
AEI (above median) × Post_2022	0.040**	0.027	-0.041*	-0.134	0.474	0.019	-0.025	-0.182***	-0.018
	[0.020]	[0.021]	[0.025]	[0.102]	[0.515]	[0.054]	[0.029]	[0.064]	[0.018]
AEI (above median)	0.002	0.033**	-0.027*	-0.101*	-1.227***	-0.047	-0.009	-0.069*	0.036***
	[0.012]	[0.013]	[0.015]	[0.057]	[0.273]	[0.030]	[0.015]	[0.038]	[0.011]
Dep. Var. Mean	0.100	0.098	0.706	4.443	14.989	12.307	0.209	3.926	0.781
Industry Category Indicators	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Obs.	8,826	8,826	14,284	14,284	14,787	14,284	12,343	13,370	16,906
Adj. R^2	0.004	0.009	0.007	0.008	0.028	0.013	0.012	0.119	0.041

This table reports the estimated effects of GenAI introduction on startup founders' previous-job characteristics and demographic characteristics using founder-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is an individual founder of startups founded between 2020 and 2024. *Expand* is an indicator for whether the founder's previous employer belongs to the top decile of the employment-growth distribution over the 12-month window surrounding the founder's job transition, while *Shrink* is an indicator for whether the previous employer belongs to the bottom decile. *≥Manager* is an indicator equal to one if the founder previously held a managerial position or above. *Seniority* measures the founder's previous-job seniority on a seven-point scale: Entry Level, Junior Level, Associate Level, Manager Level, Director Level, Executive Level, and Senior Executive Level. *Experience* is measured as years of work experience since the founder's first observed job. *Compensation* is the log of estimated total compensation in the founder's previous job, where compensation measures are obtained from Revelio Labs and are estimated using employee profiles, job titles, locations, and other worker characteristics. *Paycut* is an indicator equal to one if estimated total compensation at the startup is lower than estimated total compensation in the founder's previous job. *Education* is measured on a six-point scale ranging from high school (1) to doctorate (6). *Male* is an indicator equal to one if the founder is male. *Post_2022* equals one for startups founded in 2023–2024 and zero otherwise. *AEI (above median)* equals one for startup categories with above-median GenAI exposure. Panel A reports estimates without category controls or founded-year fixed effects. Panel B includes category indicators and founded-year fixed effects. Standard errors are obtained using a bootstrap procedure with 500 repetitions. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Internet Appendix to "Prompted to Start: How Generative AI is Transforming Entrepreneurship"

List of Appendix Figures

List of Appendix Tables

A Examples of Exposed Tasks/Occupations/Industries/Startups

Table A.1: Examples of O*NET Tasks by Share of Claude.ai Conversations

Task Description	**Share (%)**
Top 5 Tasks (Highest Shares)	
- Modify existing software to correct errors, to adapt it to new hardware, or to upgrade interfaces and improve performance.	6.65
- Diagnose, troubleshoot, and resolve hardware, software, or other network and system problems, and replace defective components when necessary.	2.82
- Modify existing software to correct errors, allow it to adapt to new hardware, or to improve its performance.	2.68
- Correct errors by making appropriate changes and rechecking the program to ensure that the desired results are produced.	2.46
- Write, update, and maintain computer programs or software packages to handle specific jobs such as tracking inventory, storing or retrieving data, or controlling other equipment.	2.40
Medium-Share Tasks	
- Advise managers on organizational policy matters such as equal employment opportunity and sexual harassment, and recommend needed changes.	0.006
- Collect information about individuals or clients, using interviews, case histories, observational techniques, and other assessment methods.	0.006
- Confer with engineers, customers, vendors, or others to discuss existing or potential electronics engineering projects or products.	0.006
- Develop and write procedures for installation, use, or troubleshooting of communications hardware or software.	0.006
- Identify or classify species of insects or allied forms, such as mites or spiders.	0.006
Low-Share Tasks	
- Teach fitness classes to improve the strength, flexibility, cardiovascular conditioning, or general fitness of participants.	0.002
- Test equipment performance, focus of lens system, diaphragm alignment, lens mounts, or film transport, using precision gauges.	0.002
- Treat wounds or superficial lacerations.	0.002
- Verify that transportation and handling procedures meet regulatory requirements.	0.002
- Write technical reports and draw charts that display statistics and data.	0.002

Table A.2: Examples of High and Low GenAI Exposure Occupations (O*NET Level)

Occupation Title	**Exposure**
High Exposure Occupations	
Computer Programmers	0.5883
Software Developers	0.4689
Software Quality Assurance Analysts and Testers	0.2869
Network and Computer Systems Administrators	0.1744
Web Developers	0.1472
Web and Digital Interface Designers	0.1472
Tutors	0.1243
Technical Writers	0.1007
Computer Systems Analysts	0.0821
Writers and Authors	0.0808
Medium Exposure Occupations	
Social and Human Service Assistants	0.0007
Farmers, Ranchers, and Other Agricultural Managers	0.0007
Transportation, Storage, and Distribution Managers	0.0006
Exercise Trainers and Group Fitness Instructors	0.0006
Private Detectives and Investigators	0.0006
Cooks, Private Household	0.0006
Environmental Engineers	0.0006
Dental Hygienists	0.0006
Industrial Engineering Technologists and Technicians	0.0006
Shipping, Receiving, and Inventory Clerks	0.0006
Low Exposure Occupations	
Laundry and Dry-Cleaning Workers	0.0001
Aircraft Structure, Surfaces, Rigging, and Systems Assemblers	0.0001
Musical Instrument Repairers and Tuners	0.0001
Roofers	0.0001
Print Binding and Finishing Workers	0.0001
Heat Treating Equipment Setters, Operators, and Tenders, Metal and Plastic	0.0001
Welders, Cutters, Solderers, and Brazers	0.0001
Commercial Pilots	0.0001
Airline Pilots, Copilots, and Flight Engineers	0.0000
Architectural and Civil Drafters	0.0000

Table A.3: Examples of High and Low GenAI Exposure Industries (4-Digit NAICS Level)

NAICS	Industry Description	Exposure
High Exposure Industries		
5132	Software Publishers	0.1207
5415	Computer Systems Design and Related Services	0.1135
3341	Computer and Peripheral Equipment Manufacturing	0.0626
3342	Communications Equipment Manufacturing	0.0520
3346	Manufacturing and Reproducing Magnetic and Optical Media	0.0455
3345	Navigational, Measuring, Electromedical, and Control Instruments Manufacturing	0.0397
4853	Taxi and Limousine Service	0.0375
5171	Wired Telecommunications Carriers	0.0368
5182	Data Processing, Hosting, and Related Services	0.0368
5192	Web Search Portals and All Other Information Services	0.0368
Medium Exposure Industries		
4411	Automobile Dealers	0.0039
5612	Facilities Support Services	0.0039
6213	Offices of Other Health Practitioners	0.0038
7212	RV (Recreational Vehicle) Parks and Recreational Camps	0.0037
4889	Other Support Activities for Transportation	0.0037
4572	Gasoline Stations with Convenience Stores	0.0037
6219	Other Ambulatory Health Care Services	0.0036
6211	Offices of Physicians	0.0034
3323	Architectural and Structural Metals Manufacturing	0.0034
3329	Other Fabricated Metal Product Manufacturing	0.0034
Low Exposure Industries		
6231	Nursing Care Facilities (Skilled Nursing Facilities)	0.0012
4922	Local Messengers and Local Delivery	0.0011
4883	Support Activities for Water Transportation	0.0011
4931	Warehousing and Storage	0.0010
4911	Postal Service	0.0010
4852	Interurban and Rural Bus Transportation	0.0010
2122	Metal Ore Mining	0.0010
4921	Couriers and Express Delivery Services	0.0009
3116	Animal Slaughtering and Processing	0.0009
2121	Coal Mining	0.0008

Table A.4: GenAI Exposure by Crunchbase Category Group

Category Group	Exposure	Category Group	Exposure
Blockchain and Cryptocurrency	0.0830	Commerce and Shopping	0.0342
AI	0.0805	Lending and Investment	0.0332
Apps	0.0743	Travel and Tourism	0.0330
Software	0.0740	Education	0.0323
Information Technology	0.0736	Administrative Services	0.0318
Data and Analytics	0.0719	Sports	0.0311
Navigation and Mapping	0.0714	Advertising	0.0307
Gaming	0.0710	Community and Lifestyle	0.0298
Privacy and Security	0.0697	Professional Services	0.0296
Payments	0.0672	Clothing and Apparel	0.0284
Mobile	0.0653	Events	0.0278
Internet Services	0.0644	Agriculture and Farming	0.0261
Platforms	0.0592	Health Care	0.0237
Hardware	0.0590	Social Impact	0.0237
Consumer Electronics	0.0565	Food and Beverage	0.0234
Science and Engineering	0.0565	Consumer Goods	0.0217
Messaging and Telecommunication	0.0556	Sustainability	0.0209
Music and Audio	0.0426	Energy	0.0201
Video	0.0412	Manufacturing	0.0194
Other	0.0405	Real Estate	0.0185
Design	0.0383	Biotechnology	0.0178
Content and Publishing	0.0375	Natural Resources	0.0167
Sales and Marketing	0.0361		
Financial Services	0.0357		
Transportation	0.0355		
Government and Military	0.0351		
Media and Entertainment	0.0349		

Table A.5: Top Ranked Startups based on GenAI Exposure

Company	Short Description
SuperDuperAi	SaaS for Generative AI.
Sapix Lab, Inc.	Redefines entertainment products through the power of Generative AI.
CRACKENAGI, Inc.	IT system testing and evaluation.
Comfy Deploy, Inc.	SaaS company providing production-ready APIs for generative AI applications.
Experio Labs, Inc.	Knowledge management platform using AI and machine learning.
Ablt AI, Inc.	Enables rapid deployment of Generative AI capabilities for any business.
Cloobot Techlabs Pvt Ltd.	Builds implementation-as-a-service solutions for enterprise CRM/ERP systems.
Langgenius, Inc.	Dify is an open-source LLM app development platform integrating Backend-as-a-Service and LLMOps.
Infinitform, Inc.	AI-powered design tool optimizing CAD parts for real-world manufacturing processes.
Lilac Voice AI	Conversational AI agent for drive-thru order taking.
heyLevi	Zero-prompt AI software that creates brand identity and positioning for businesses.
TalkStack AI	No-code GenAI customer experience platform for enterprises without AI teams.
Simprobot, Inc.	On-premise AI solution for corporations.
Statra LLC	Analytics platform addressing reporting challenges in healthcare organizations.
LOMA (US)	Marketing platform to scale and validate local marketing for multi-unit brands.
Tabnam, Inc.	Greptile provides AI-driven code review and development tool integration.
Preloop, Inc.	Platform to deploy ML models in hours instead of weeks.
Anote	Human-centered AI platform combining generative models with human feedback on edge cases.
Goml	Enterprise solutions provider in data engineering, MLOps, and LLM applications.
Pull Data, Inc.	Machine-learning-as-a-service solution for capturing and analyzing vehicle data.

B Identifying AI-Enabled and AI-Native Startups Using LLM Pipelines

We construct two startup-level measures that capture related but conceptually distinct dimensions of AI exposure. The first measure, `AI_enabled`, identifies startups that use AI technologies or AI tools in their internal business operations in a substantive and verifiable way. The second measure, `AI_native`, identifies startups for which AI is central to the firm's product or service architecture, value proposition, production technology, or business model. We therefore treat AI enablement as an internal adoption measure and AI nativeness as a product- and business-model measure. We implement two separate LLM-based feature engineering pipelines to construct these variables. Both pipelines are designed to prioritize precision, traceability, and reproducibility. In particular, positive classifications must be supported by citable external evidence, and the prompts explicitly prohibit unsupported inference from vague or promotional language.

B.1 Startup AI-Enablement Identification

We identify AI-enabled startups using a two-stage agentic research pipeline that combines lightweight online signal screening with deeper web-based evidence retrieval. The purpose of the first stage is to determine whether a startup has enough online information to justify a more costly evidence-gathering step. The purpose of the second stage is to retrieve, evaluate, and record specific evidence of internal AI adoption.

In *Stage 1*, we assess whether each startup is sufficiently researchable online. Starting from the Crunchbase startup profile, we run a website health check and query the Tavily Crawler API (`https://www.tavily.com/`) to obtain web search results for the startup. These inputs are evaluated by an LLM-based classifier, using the `gpt-5 nano` model on March 10, 2026. The classifier assigns a `research_priority_score` from 0 to 5 based on the number of results, source diversity, snippet relevance, and confidence that the retrieved results correspond to the focal startup. Startups with weak online signals are not escalated. Startups with scores of 4 or 5 are passed to Stage 2.

In *Stage 2*, escalated startups are processed using the Perplexity Agent API in `deep-research` mode, using the `gpt-5.2` model on March 19, 2026. This stage combines a frontier LLM with live web search and page-fetching tools. For each startup, the agent performs iterative search-and-reasoning steps, refines search queries, and aggregates evidence across multiple sources. The agent is configured with bounded retrieval and reasoning parameters, including limits on search iterations, page retrieval, and token usage, to promote consistent processing across startups. This design bases classification on actively retrieved and synthesized evidence rather than on static firm descriptions or one-shot prompting.

The AI-enabled prompt encodes several research-validity rules directly in the system instructions. First, *no source implies no finding*: every positive claim must be supported by a citable external URL. Second, the agent is instructed to report only what the source explicitly states and not to infer internal adoption from general references to AI, automation, or innovation. Third, findings should be tool-specific whenever possible. Generic descriptions such as "AI coding tool" are not accepted when the source identifies a specific product such as `GitHub Copilot`, `Cursor`, or `DALL-E`. Fourth, the prompt imposes a *USE versus SELL* guardrail: a startup building or selling AI products is not, by itself, evidence that the startup uses AI internally. Internal adoption must be supported by separate evidence about the firm's own operational use of AI.

A startup is classified as AI-enabled if the Stage 2 dossier contains verifiable evidence that the firm uses AI tools in at least one internal business function or operational workflow. These functions include engineering, marketing, operations, product and design, customer service, sales, human resources, and finance. All Stage 2 responses are returned in machine-readable JSON. For each finding, the output records the identified AI tool, business function, use case, evidence description, source URL, source type, and model metadata. Evidence is drawn from several source types, including company blogs, job postings, podcasts, videos, interviews, company websites, news articles, and press releases.

B.2 AI-Native Startup Taxonomy Classification

We classify AI-native startups using a two-stage Tavily-enriched LLM pipeline that maps each startup into a preset taxonomy. This pipeline focuses on whether AI is central to the firm's core product, service, production technology, or business model, rather than on whether the firm uses AI tools internally.

In *Stage 1*, we construct the evidence base used for classification. The input combines structured Crunchbase information about the startup, evidence from the startup's website, and web-crawl text retrieved through the Tavily Search API. This stage is designed to collect startup-specific information that can help the classifier distinguish core product and business-model features from generic AI references or peripheral uses of AI.

In *Stage 2*, the LLM classifier maps each startup into one mutually exclusive category in the taxonomy. The classifier is implemented using the `gpt-5.4 nano` model on May 5, 2026 and uses the Stage 1 evidence base as input. The taxonomy separates AI-native and non-AI-native firms and then assigns firms to finer subcategories within each group.

The AI-native taxonomy contains seven subcategories. *Foundation-layer firms* develop core AI infrastructure, models, or data layers. *Applied vertical AI firms* embed AI into products designed for specific industries or workflows. *AI-native tooling firms* provide developer or enterprise tools for building, deploying, monitoring, or governing AI systems. *Autonomous-agent firms* build systems that can execute multi-step tasks with limited human intervention. *Generative content platforms* produce text, image, audio, video, code, or other synthetic content. *Thick LLM integrators* combine foundation models with substantial proprietary workflow, data, or product layers. *Thin LLM wrappers* provide relatively narrow interfaces built primarily around third-party LLM APIs. Firms that are not classified as AI-native are assigned to one of three non-AI-native categories: traditional tech, non-tech, or AI-augmented.

This taxonomy allows us to distinguish firms whose business models depend directly on AI from firms that merely mention AI, sell conventional software, or use AI as a peripheral feature. The classifier returns a binary `AI_native` indicator, a mutually exclusive taxonomy category, and a confidence score. For startups classified as AI-native, the classifier also returns a RAD score, which measures reliance on third-party AI APIs. We classify this reliance as `RAD-H`, `RAD-M`, or `RAD-L`, corresponding to high, medium, or low reliance on external AI APIs. A high-RAD firm depends structurally on third-party models or APIs for its core product functionality. A medium-RAD firm uses external AI services but also has meaningful proprietary workflow, data, or infrastructure layers. A low-RAD firm relies mainly on in-house models, proprietary infrastructure, or AI capabilities that are less directly tied to third-party APIs. The output therefore supports both the construction of the binary `AI_native` indicator and heterogeneity analyses based on finer taxonomy assignments and RAD scores.

B.3 Summary Statistics

Among the approximately 44,000 startups processed through the AI-enabled and AI-native pipelines, about 21% and 48%, respectively, meet the conditions to enter the advanced stage of the relevant pipeline. As shown in Figure B.1, the share of startups with sufficient online information to enter the advanced stage varies over time and differs between the AI-enabled pipeline, where the advanced stage is Stage 2 deep research, and the AI-native pipeline, where the advanced stage is the Stage 2 LLM classifier using the Tavily-enriched evidence base. This coverage measure is important because both pipelines rely on verifiable online evidence. Firms without sufficient retrievable information are not escalated and are therefore less likely to receive a positive classification based on external evidence.

For the AI-enabled pipeline, among the 9,400 firms eligible for Stage 2 deep research, approximately 1,250 firms, or 13%, yield at least one verified AI adoption finding. The pipeline produces 2,062 total verified findings. In terms of tool identification, approximately 75% of findings reference a specific GenAI product, while 25% document AI usage without naming a particular tool. Because a single finding may identify multiple tools, the distribution of named tools is reported at the tool-mention level. Among named-tool mentions, ChatGPT/OpenAI accounts for the largest share, with 850 mentions, or approximately 49%. Claude/Anthropic follows with 393 mentions, or 22%; GitHub Copilot with 209 mentions, or 12%; Cursor with 196 mentions, or 11%; and Gemini/Google AI with 98 mentions, or 6%. This distribution indicates that documented internal adoption is concentrated in general-purpose LLMs and coding assistants.

Evidence sources are diverse but concentrated in a few high-signal channels. Company blogs account for 36% of findings and job postings account for 25%. Together, these two source types represent more than 60% of all evidence. Additional sources include podcasts, interviews, videos, press coverage, company websites, and press releases, which provide complementary but less frequent evidence.

Table B.1 provides representative examples of firm-level AI adoption findings. The table illustrates the structure of the recorded evidence: each example identifies the startup, internal business function, AI tool, source type, and the specific evidence supporting the classification. All examples satisfy the pipeline's evidence criteria: each observation is supported by a verifiable source URL, identifies the tool when available, and documents usage in a concrete business context.

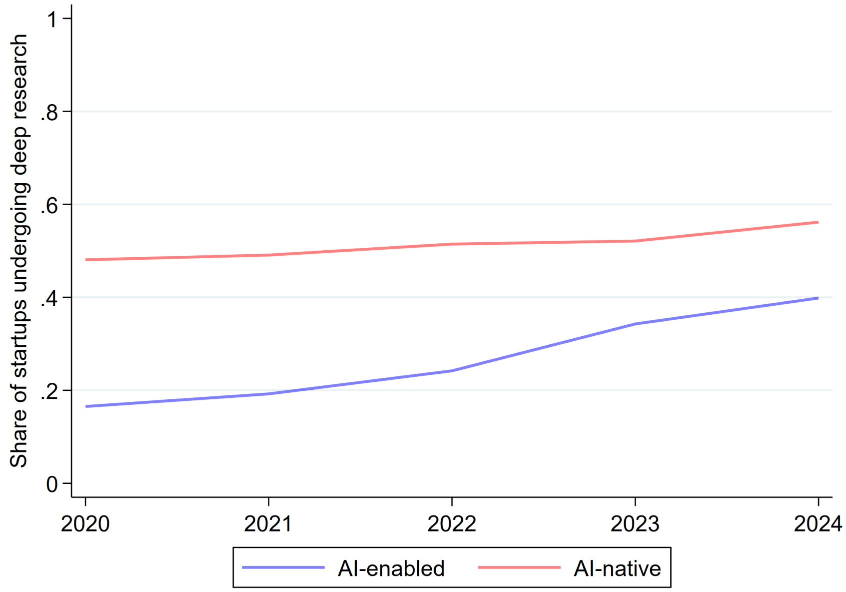
For the AI-native pipeline, the Tavily-enriched classification sample contains 21,510 U.S. startups with sufficient website evidence. Among them, 3,348 startups, or 15.6%, are classified as AI-native. The largest AI-native category is applied vertical AI, with 1,908 startups. This is followed by thick LLM integrators with 639 startups, AI-native tooling firms with 334 startups, autonomous-agent firms with 322 startups, generative content platforms with 59 startups, foundation-layer firms with 45 startups, and thin LLM wrappers with 41 startups. Among non-AI-native firms, the vast majority are classified as traditional tech firms, with 18,056 startups; smaller groups are classified as non-tech or AI-augmented.

The RAD score measures the extent to which an AI-native startup depends on third-party AI APIs for its core product functionality. Among AI-native firms, 1,465 are classified as `RAD-H`, indicating high structural reliance on external AI models or APIs. Another 1,709

Table B.1: Representative AI-Enablement Findings

Startup	Function	Tool(s)	Source type	Evidence summary
Sawmills	Engineering	Claude Opus; GPT-01/Codex; Haiku 4.5	Company blog	The CTO describes adopting agentic coding tools internally, including Claude Opus for planning, Haiku 4.5 for execution, and GPT-based models for debugging and architectural tasks.
Infinite Renewals	Marketing	Jasper	Podcast transcript	The founder states that the firm uses Jasper for writing and content creation tasks, providing evidence of GenAI adoption in marketing workflows.
Steinert Analytics	Operations	Claude Pro (Sonnet 4.5)	Company blog	A company blog post on financial analysis workflows recommends and references Claude Pro for structuring calculations and drafting outputs, indicating GenAI usage in operational decision support.
Daisy	Customer service	ChatGPT	Company blog	A blog post on remote support describes the use of ChatGPT to assist technicians in troubleshooting and responding to customer queries.
Nitrility	Sales	ChatGPT (GPT-4)	Interview	The founder describes using ChatGPT to draft investor communications and fundraising messages, providing evidence of GenAI usage in sales and external communication.
Blue Sky Robotics	Human resources	ChatGPT (GPT-4.5)	Podcast transcript	A podcast transcript documents the CEO using ChatGPT for role-playing and preparing responses to staff communication scenarios.

Figure B.1: Startup Coverage in the Advanced Stage of the LLM Pipelines

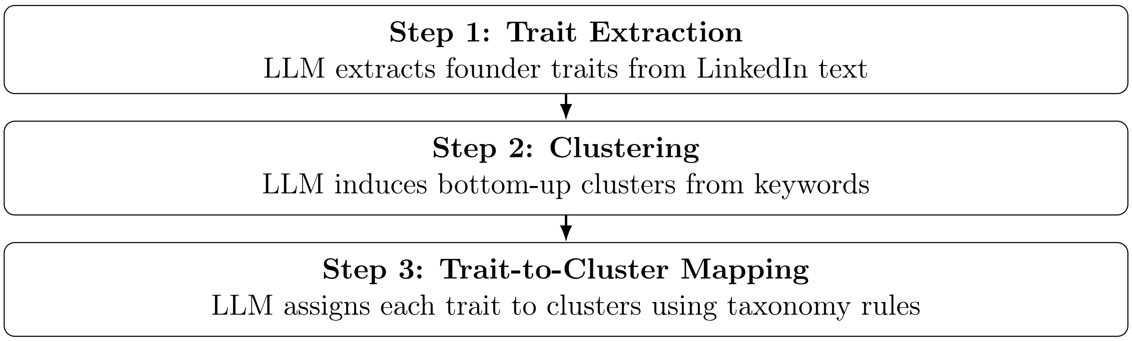


This figure plots the share of startups with sufficient online information to enter the advanced stage of the LLM pipelines from 2020 to 2024. The series are shown separately for the AI-enabled and AI-native classification pipelines. In the AI-enabled pipeline, the advanced stage is Stage 2 deep research. In the AI-native pipeline, the advanced stage is the Stage 2 LLM classifier that uses the Tavily-enriched evidence base constructed in Stage 1.

are classified as `RAD-M`, indicating meaningful use of external AI services combined with proprietary workflow, data, or product layers. Finally, 174 are classified as `RAD-L`, indicating relatively low dependence on third-party AI APIs. Thus, most AI-native startups in the sample rely at least partly on external AI infrastructure, but a large share also adds nontrivial proprietary layers on top of it. The AI-native rate is substantially higher among firms founded in the AI era, reaching 42.8% for startups founded in 2023 or later.

C LLM-Assisted Founder Characteristics Comparison

We use a three-step LLM-assisted pipeline to compare founder characteristics for companies established pre- versus post-GenAI.



Step 1: Trait extraction. After matching founders from Crunchbase to LinkedIn profiles, we collect free-text entries that describe each entrepreneur, including profile headline/title, summary ("About"), and job descriptions, restricted to the founding year and the year prior. In our sample, 14,062 of 14,149 pre-GenAI founders and 1,857 of 1,869 post-GenAI founders have non-empty text. We then prompt an LLM to extract concise traits (keywords/attributes) that characterize the person. We also require the LLM to differentiate whether a trait is directly observed or logically inferred; in either cases, it also needs to provide the text quote based on which it extracted the trait. This yields over 55k (pre-GenAI) and 11k (post-GenAI) unique traits. The model used in this step is Gemini-3-Flash.

Step 2: Bottom-up cluster induction. We pool traits across both periods ($\sim$ 61k unique traits) and ask an LLM to induce a taxonomy in a strictly bottom-up manner: clusters are formed solely from semantic similarity within the observed trait list, without imposing an external framework. We provide a sequence of tasks for the LLM to follow:

- Normalize keywords by standardizing casing, punctuation, plurals, and hyphenation, and recording near-duplicate aliases.

- Induce bottom-up clusters from the terms present with minimal overlap, marking any necessary multi-membership as multi-label candidate and keeping Misc/Other small.

- Consolidate into a stable final cluster set by merging redundancies and splitting only when clear subgroups exist.
- Generate mapping rules for each cluster using strong inclusion cues and clear exclusion cues to avoid nearest-neighbor confusion.
- Add post-hoc academic annotations by attaching relevant organizational behavior or entrepreneurship lenses with citations when confident, or search terms when not.

The model used is GPT-5.2 (thinking mode), and it produces 18 clusters in this unsupervised step.

Step 3: Trait-to-cluster mapping. Finally, we prompt an LLM to assign each trait to 0-3 cluster labels using lexical anchors and boundary rules derived from step 2. Multi-label assignments are allowed when a trait clearly spans multiple domains (e.g., "AI governance" maps to both AI/ML and compliance/security). The model used in this step is GPT-5-mini.

Table C.1: Cluster Description and Representative Examples (1/2)

ID	Cluster name	Inclusion rules (high-signal cues)	Examples
1	Venture roles & startup-stage markers	founder/co-founder; serial entrepreneur; startup CEO/CTO; first employee; accelerator/incubator as role-stage identity	"Co-Founder", "CEO", "Chief Product Officer"
2	Outcomes, traction & scale metrics	ARR/MRR/AUM; revenue/retention/subscribers; NPS; uptime; exit/acquired as outcomes; large numeric/currency growth indicators	"High-growth track record", "Successful exit"
3	Capital, investing & deal mechanics	angel/VC/seed; PE; underwriting; financing structures; capital stack; M&A; portfolio/asset management; exchanges as deal mechanics	"VC-backed", "Fundraising capability"
4	Go-to-market (marketing/sales/growth/ partnerships)	GTM; marketing/sales; demand gen; ABM; channels; partnerships/alliances; ecommerce selling; marketplace growth mechanics	"User Growth", "Marketing Leadership"
5	Product building & "0–1" creation	product manager/owner; product development; launch/MVP/build-from-scratch; 0–1 / $0 \to 1$ phrasing; builder language tied to creation	"Product development", "Full-cycle Product Experience"
6	Software engineering (apps/web/platforms/ frameworks)	developer roles; programming languages; frameworks/toolkits; app/web development; platform/API implementation terms	"C++", "Mobile App Developer", "Web frontend developer"
7	Data/analytics & experimentation	analytics/BI; dashboards; measurement/attribution; A/B testing; experiment design; insights; regression/statistics phrasing	"Data-driven", "Data Analytics Expert"
8	AI/ML & GenAI (agents, safety, governance)	AI/ML; LLMs; deep learning; computer vision; agents; fine-tuning; AI safety/governance explicitly	"AI/ML focus", "Full-stack AI Developer"
9	Cloud/infra/DevOps & reliability/observability	cloud providers; migrations; infra; observability/monitoring; scalability/reliability; DevOps/SRE terms	"Cloud Infrastructure Developer", "DevOps Specialist"

Table C.2: Cluster Description and Representative Examples (2/2)

ID	Cluster name	Inclusion rules (high-signal cues)	Examples
10	Security, risk, privacy & compliance	security/privacy; risk management; audit/compliance; fraud/AML; governance-as-control/assurance; enforcement	"Compliance Specialist", "Cybersecurity Focus"
11	Design/creative/content production	3D modeling/rendering; animation; Adobe suite; media production; visual/creative craft terms	"Copywriter", "UX Designer", "User-Centric Design"
12	Operations, service delivery & logistics	operations; process; implementation; service delivery; fulfillment/3PL; support systems; logistics	"Inventory Management", "Operational Efficiency", "Process Optimizer"
13	Academic/research/ clinical orientation	professor/researcher; PhD; clinical practice; laboratory; university affiliation as role identity	"Principal Scientist", "Professor", "Researcher"
14	Credentials, memberships, awards & public recognition	certified/certification; fellow/member; awards/honors; rankings; patents as recognition; alumni as credential	"30 Under 30 Honoree", "EY Entrepreneur of the Year"
15	Values, advocacy & impact orientation	advocacy; accessibility; equity/inclusion; affordability; sustainability; ethical/mission-driven for-good language	"Sustainability Focus", "Mission-driven"
16	Personal dispositions & working style	trait adjectives (analytical, adaptable, ambitious, accountable, action-oriented, etc.)	"Client-Centric", "Tech savvy"
17	Industry/vertical domain tags	explicit industry/vertical tags. Sub-cluster tags: 17a Healthcare & life sciences, 17b Financial services / fintech / insurance, 17c Real estate / proptech / housing, 17d Aerospace / defense / industrial / manufacturing, 17e Energy / climate / cleantech, 17f Agriculture / food / agtech, 17g Media / entertainment / gaming / creator economy	"Biomedical expertise", "Commodities trader", "Clean energy focus", "Licensed food manufacturer"
18	Entities/noise/ unclassifiable tokens	opaque proper nouns; very short tokens; numeric-only items; insufficient semantics to match other cues	

D Additional Tables and Figures

Table D.1: Summary Statistics (QWI)

Variable	N	Mean	Std. Dev.	Median	p1	p25	p75	p99
Employment (age=0-1)	5,218	18,434	67,520	3,893	22	1,006	12,491	140,538
Employment (age>1)	20,890	114,898	406,173	11,158	29	2,292	68,237	1,735,776
Earnings (age=0-1)	5,211	4,624	2,363	4,158	1,463	3,019	5,647	15,104
Earnings (age>1)	20,880	5,249	2,736	4,600	1,660	3,387	6,312	15,282

This table reports summary statistics on employment and wages at the year–quarter–industry level for firms in different age groups, using data from QWI.

Table D.2: GenAI Introduction and the Geographic Distribution of Entry: Channels
Alternative Measures of Labor Specialization and VC hubs

Panel A: Relaxed Labor Specialization Requirements

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
	non-specialized	specialized	non-specialized	specialized
AEI (above median) × Post_2022	0.091**	0.046	0.204**	0.055
	[0.037]	[0.037]	[0.088]	[0.061]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.477	0.574	0.887	0.851

Panel B: Relaxed Financing Requirements

	(1)	(2)	(3)	(4)
Dep. Var.	I (firm entry)		# firm entry (log)	
	non VC hubs	VC hubs	non VC hubs	VC hubs
AEI (above median) × Post_2022	0.107***	-0.01	0.216**	-0.035
	[0.036]	[0.038]	[0.087]	[0.063]
Industry FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430
Adj R2	0.491	0.58	0.89	0.81

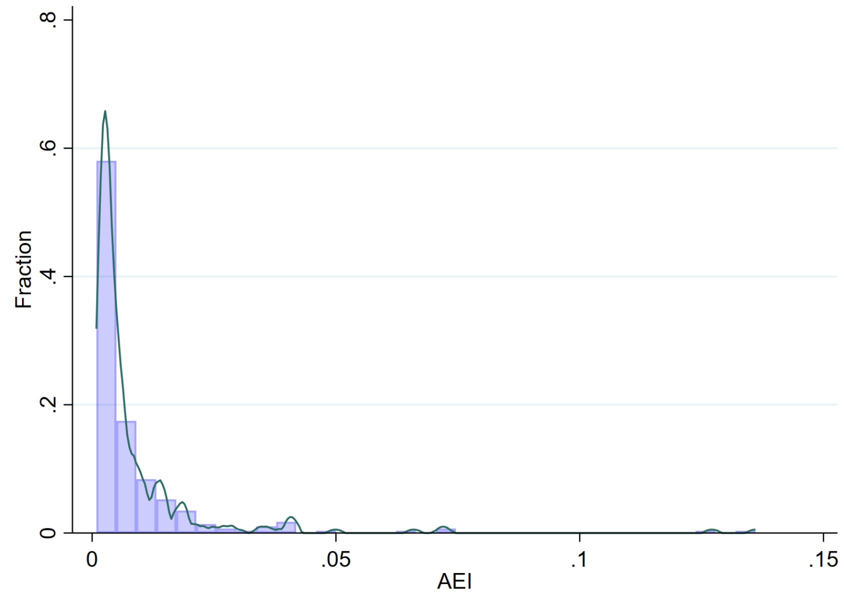
This table examines two channels through which the introduction of GenAI may have contributed to more geographically dispersed firm entry with alternative measures of labor specialization and VC hubs, using establishment-level data aggregated from the InfoGroup firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. Panel A examines the labor specialization channel using a broader geographic unit. Each combined statistical area (CSA) is classified as having *specialized* or *non-specialized* labor in a given industry based on the pre-GenAI distribution of employment. For each industry, the share of local employment in that industry relative to total CSA employment is computed, and CSAs in the top decile of this distribution are defined as specialized. Panel B tests the financing channel using an alternative definition of VC hubs. Venture capital intensity is measured using data from Crunchbase on all U.S.-based investors. The total dollar value of investments made by investors located in each county is aggregated, and the top ten counties in this distribution are classified as VC hubs, with all others treated as non-hubs. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Table D.3: Effect of GenAI Introduction on the Size Composition of New Firms
Absolute Size Thresholds

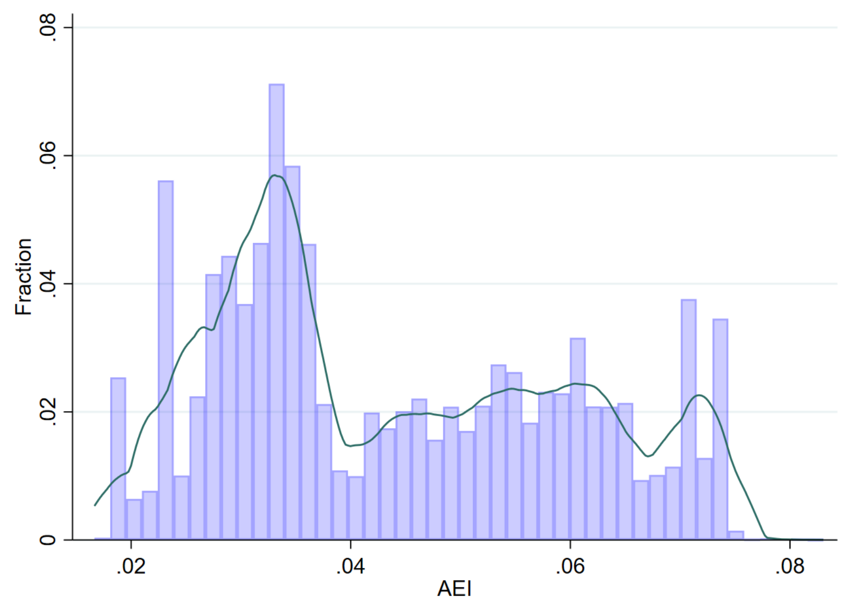
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	# firm entry (log)					
	≤ 3 employees	> 3 employees	≤ 5 employees	> 5 employees	≤ 10 employees	> 10 employees
AEI (above median) × Post_2022	0.183**	0.031	0.174**	0.029	0.180**	0.06
	[0.075]	[0.087]	[0.079]	[0.085]	[0.082]	[0.083]
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Obs	1,430	1,430	1,430	1,430	1,430	1,430
Adj R2	0.873	0.877	0.876	0.853	0.877	0.769

This table reports the estimated effects of GenAI introduction on the size distribution of new firm entries, using establishment-level data aggregated from the *InfoGroup* firm panel. The unit of observation is a four-digit NAICS industry–year cell. The sample includes newly established, single-location firms founded between 2020 and 2024. The dependent variable is the logarithm of one plus the number of new firm entries (*# firm entry (log)*) in each industry–year cell, separately for firms with initial employment below or above the pre-treatment industry median. Columns (1)–(6) report estimates for firms with initial employment below versus above absolute size thresholds of 3, 5, and 10 employees, respectively. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for industries whose exposure is above median, and zero otherwise. Industry (four-digit NAICS) and year fixed effects are included in all specifications. Standard errors are clustered at four-digit NAICS level. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

Figure D.1: GenAI Exposure Distribution at the Industry and Startup Levels



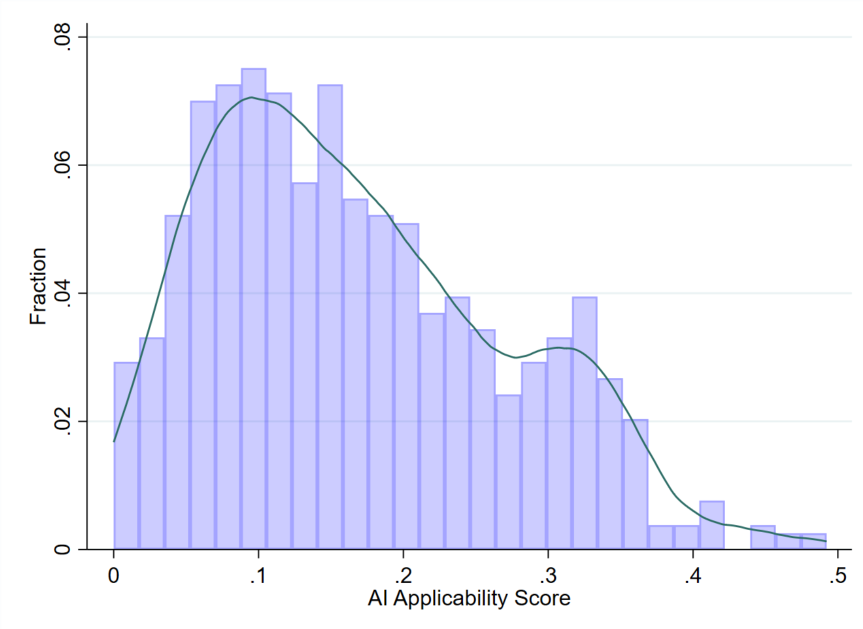
(a) Industry (NAICS 4-digit)



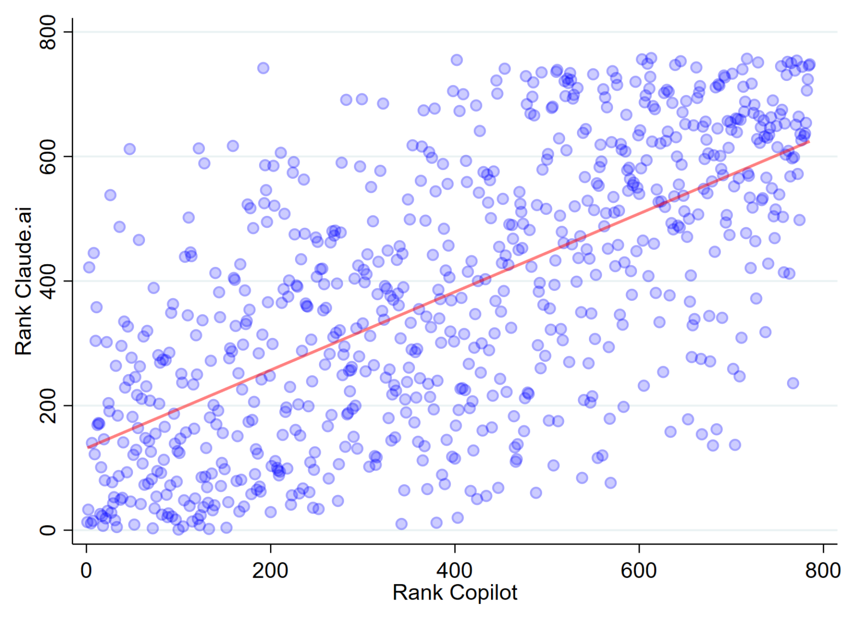
(b) Startup

This figure illustrates the distribution of exposure to GenAI based on the Claude.ai conversation data. Panel (a) presents the histogram and kernel density of GenAI exposure aggregated at the four-digit NAICS industry level, while Panel (b) displays the corresponding distribution across startups.

Figure D.2: GenAI Exposure Distribution Based on Microsoft Bing Copilot



(a) Distribution Based on Bing Copilot (Occupation Level)



(b) Claude.ai Rank vs. Bing Copilot Rank (Occupation Level)

This figure compares exposure to GenAI derived from Microsoft Bing Copilot data with the baseline measure based on Claude.ai conversations. Panel (a) presents the kernel density distribution of the Copilot-based AI Applicability Score across roughly 750 occupations. Panel (b) displays the rank correlation between the Claude.ai exposure measure and the Copilot-based applicability score at the occupation level.

Table D.4: Effects of GenAI Introduction on Startup Founder Traits: Observed and Inferred Trait Counts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. Var.	Capital/ Investing	AI/ML GenAI	Product Manager/ Development	Marketing/ Sales	Operations/ Logistics	Capital/ Investing	AI/ML GenAI	Product Manager/ Development	Marketing/ Sales	Operations/ Logistics
AEI (above median) × Post_2022	-0.152**	0.258***	-0.061	-0.083	-0.118	-0.148**	0.224***	-0.067	-0.070	-0.107
	[0.061]	[0.038]	[0.042]	[0.089]	[0.081]	[0.061]	[0.038]	[0.042]	[0.089]	[0.082]
Post_2022	0.133**	0.043*	0.058*	0.092	0.075					
	[0.053]	[0.025]	[0.034]	[0.075]	[0.069]					
AEI (above median)	-0.162***	0.216***	0.305***	0.036	-0.183***	-0.186***	0.010	0.280***	0.114***	-0.099***
	[0.019]	[0.010]	[0.013]	[0.024]	[0.025]	[0.022]	[0.013]	[0.019]	[0.034]	[0.032]
Total observed and inferred traits	0.004***	0.001***	0.006***	0.016***	0.058***	0.004***	0.001**	0.006***	0.016***	0.058***
	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]	[0.000]	[0.000]	[0.000]	[0.001]	[0.001]
Dep. Var. Mean	0.475	0.229	0.463	0.782	1.321	0.475	0.229	0.463	0.782	1.321
Industry Category Indicators	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES
Year FE	NO	NO	NO	NO	NO	YES	YES	YES	YES	YES
Obs	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154	15,154
Adj R2	0.011	0.052	0.043	0.032	0.309	0.012	0.080	0.043	0.034	0.310

This table reports the estimated effects of GenAI introduction on observed and inferred founder trait counts, using founder-level data from the Crunchbase–LinkedIn matched sample. The unit of observation is an individual founder of startups founded between 2020 and 2024. The dependent variables are counts of observed and inferred traits in five trait clusters defined in the appendix, Appendix C: Cluster 3, capital, investing, and deal mechanics; Cluster 8, AI/ML and GenAI; Cluster 5, product building and "0–1" creation; Cluster 4, go-to-market, including marketing, sales, growth, and partnerships; and Cluster 12, operations, service delivery, and logistics. *Post_2022* is a dummy that equals one for the post-GenAI period (2023–2024) and zero otherwise. *AEI (above median)* is a dummy equal to one for firms operating in category groups with above-median GenAI exposure, and zero otherwise. All columns control for the total number of observed and inferred traits extracted for the founder across the 16 trait groups. Columns (1)–(5) report estimates without category-group controls or year fixed effects, while Columns (6)–(10) include both sets of controls. Standard errors are obtained via a bootstrap procedure with 500 repetitions. * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.