Prompted to Start: How Generative AI is Transforming Entrepreneurship*

Jan Bena† Bo Bian‡ Mariassunta Giannetti§

March 16, 2026

Abstract

Generative AI's impact on the economy is unclear, with concerns about worker displacement competing against predictions of productivity-driven growth. We show that entrepreneurial entry is a critical margin through which GenAI reshapes labor demand. Following the diffusion of the new technology, industries with higher task-level exposure to GenAI experience 20% more startup formation than less-exposed industries, and an increasing proportion of startups offer new AI services and products. This entry is geographically dispersed, extending beyond traditional innovation hubs into regions with limited venture capital and thin specialized labor markets. While individual startups in exposed industries enter at a smaller scale, aggregate entry generates net employment and wage growth at the industry level. Crucially, GenAI changes who becomes an entrepreneur: founding teams are increasingly drawn from occupations with high exposure to GenAI, suggesting that the technology enables new combinations of human capital rather than simply replacing workers. Our findings reveal that business dynamism partially offsets the labor-demand contractions GenAI produces among incumbent firms.

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

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

*We thank the participants in the Bank of Canada, Collegio Carlo Alberto-Università di Torino, and HEC Montreal Conference on Policy Challenges in a Changing Global Economy 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 and complementarity with organizational and intangible investments.[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. This raises concerns about labor displacement that may extend even to well-paid occupations (Acemoglu et al., 2022; Felten et al., 2023). Existing evidence remains inconclusive, in part because the adoption and diffusion of new technologies are typically gradual.

We contribute to this debate by examining how GenAI affects entrepreneurial entry and the characteristics of new ventures. Startups provide a natural setting to study GenAI's early economic effects for three reasons. First, young firms adopt new technologies more rapidly than incumbents, offering a leading indicator of broader economic impacts. Second, startups play a central role in job creation and reallocation (Haltiwanger et al., 2013; Decker et al., 2016). Third, the organizational choices of new ventures, including team composition, capital intensity, and geographic location, reveal how entrepreneurs adapt production to exploit new technological opportunities. To the best of our knowledge, we are the first to systematically document how GenAI is reshaping firm entry and startup characteristics across industries and to show that labor demand in new ventures differs markedly from that in established firms.

GenAI can influence startup formation through three non-mutually exclusive channels. First, by expanding the production possibility frontier, GenAI enables new products and services that were previously technically or economically infeasible. Startups whose core offerings rely on GenAI capabilities—such as AI-powered design tools, automated content generation, or conversational interfaces—represent pure technology-driven entry. Second,

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

even when GenAI is peripheral to a firm's value proposition, it may reduce fixed entry costs by automating non-core tasks such as prototyping, coding, marketing, and customer support. Thanks to lower resource requirements, GenAI enables startups to increase output without proportional growth in employment. This reduces reliance on specialized labor and external finance and may weaken the geographic concentration of entrepreneurship in venture capital centers (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. Textual analysis of business descriptions reveals that entrepreneurial

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

entry increases because of two channels: (i) startups automating existing processes using GenAI, and (ii) "AI-native" ventures offering new products and services built on GenAI capabilities. The prevalence of AI-native startups suggests GenAI is creating new markets rather than simply reducing costs in existing ones.

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 reducing barriers to entry by lowering fixed costs and enabling leaner operations.

Third, while individual startups in high-exposure industries enter at smaller scale, aggregate entry generates net employment growth. Higher entry rates more than offset smaller firm size, resulting in a 5% increase in employment by new firms in high-exposure 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. Founders in high-exposure industries are significantly more likely to have prior work experience in GenAI-exposed occupations, enabling them to leverage domain expertise while using GenAI to augment their productivity. This contrasts sharply with established firms, where employment in GenAI-exposed occupations is declining (Acemoglu et al., 2022). Our evidence indicates that GenAI complements rather than substitutes for workers in these occupations when they transition to entrepreneurship, suggesting new ventures recombine human capital differently than incumbents.

These findings carry important implications for understanding GenAI's economic impact. The entrepreneurship margin reveals forces absent in studies of incumbent firms. While established companies may shed workers in GenAI-exposed occupations to capture cost savings, startups hire precisely these workers to exploit new opportunities the technology creates. Business dynamism can thus mediate the labor-demand contractions GenAI produces among incumbents. Moreover, the geographic dispersion of entry and the reduced educational requirements of founders suggest GenAI may partially democratize entrepreneurship, though we find mixed evidence on demographic diversity.

Our paper contributes to several literatures. A growing body of work examines GenAI's labor-market implications, documenting declining demand for GenAI-exposed occupations in established firms (Acemoglu et al., 2022; Brynjolfsson et al., 2018) alongside productivity gains for less-exposed workers (Hampole et al., 2025). We examine entry patterns across all industries and document how startups organize production around GenAI.

A second 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 and exposure measures. Section 3 presents our main results on entry patterns and geographic dispersion. Section 4 analyzes startup characteristics, including size and employment composition. Section 5 discusses startups' business models and their implications for long-run productivity growth. Section 6 examines the characteristics of founding teams. Section 7 presents several robustness tests. Section 7 concludes.

2 Data and Measurement

We use a variety of data sources to measure entrepreneurial entry, firm dynamics, industry exposure to AI, 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 EIN application form includes the applicant's name and address, the business name and address, the business start date, the type of business entity, the principal industry, and the planned date of initial wage payments (if applicable).

The Census Bureau publishes monthly the number of EINs applications submitted to the IRS at the two-digit NAICS level.[3] 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, rather than merely registration activity, and are commonly used to provide early signals of nascent entrepreneurship (Dinlersoz et al., 2021).

For our purposes, this dataset offers two key advantages. First, entrepreneurs typically apply for EINs at the very beginning of the firm-creation process, well before the firm becomes visible in commercial databases such as Crunchbase or InfoGroup. This makes the BFS a

[3]The two-digit sectors include agriculture (11), mining (21), utilities (22), construction (23), manufacturing (MNF), wholesale trade (42), retail trade (RET), transportation and warehousing (TW), information (51), finance and insurance (52), real estate (53), professional services (54), management (55), administrative services (56), education (61), health care (62), arts and entertainment (71), accommodation and food services (72), other services (81), and government (92).

timely measure of new business intent and formation. Second, the monthly frequency enables us to track the dynamics of new business entry in real-time, around the introduction and diffusion of GenAI technologies. However, the BFS data also has important limitations. Industry information is available only at the two-digit NAICS level, meaning that measures of GenAI exposure are necessarily coarse and potentially noisy. Additionally, the industry-level aggregation of the data precludes the analysis of firm-level characteristics, such as size, team composition, or occupational structure. For these reasons, we use the BFS primarily to provide descriptive and motivating evidence, and then turn to more granular datasets at the firm and founder levels to conduct complementary analyses of organizational and workforce dynamics.

To capture industry dynamics at a higher level of granularity, we use InfoGroup (now operating under the name *Data Axle*), which provides a comprehensive firm-level database that serves as a private-sector alternative to official Census business registries.[4] 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 has been widely used in academic research (Acharya et al., 2023; Borisov et al., 2021; Denes et al., 2023). It tracks the universe of U.S. establishments, including both newly registered and existing firms, and contains information on founding year, location, industry classification (NAICS), and employment. Infogroup's granularity enables the analysis of new firm creation, size distribution, and industry composition at fine levels of industry aggregation (e.g., four-digit NAICS).

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 test, we distinguish between very young firms (ages 0–1) and the rest of the

[4] Unlike Census data, which are typically released with multi-year lags, InfoGroup offers a timelier, continuously updated view of new business formation.

companies to evaluate the employment effects of startups. Table E.1 provides summary statistics.

2.2 Startups and Founding Teams' Characteristics

Finally, to study how GenAI is altering the types of individuals who become founders and the occupational mix in new firms, we construct an employer-employee-matched database. 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. We obtain information on Crunchbase's proprietary industry classifications for startups and measure their employment at the end of the founding year.[5] We exclude a small number of outlier observations (approximately 0.5% of the sample) corresponding to firms with more than 100 employees at the end of the founding year, as these are likely to represent data misclassifications, subsidiary relaunches, or unusually large corporate spinouts rather than genuine early-stage companies.

To obtain further information on the startups and their founding teams, we link firm information from Crunchbase to employee-level data from LinkedIn, which we obtained through Revelio Lab, using the firm-specific LinkedIn URLs provided in Crunchbase. Next, 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.[6]

We further identify the founders of each startup using the founder records provided in Crunchbase. For each founder, we extract individual-level information—including profes-

[5] Approximately 40% of Crunchbase entries lack founding date information and are therefore excluded from the sample. Moreover, the recorded founding dates are often imprecise and appear to have been backfilled with the first day of the year. For this reason, we focus on employment measured at the end of the founding year, which provides a comparable benchmark across startups.

[6] The eight-digit occupational code extends the official Standard Occupational Classification (SOC) six-digit system by adding two decimal digits at the end. The extra digits distinguish detailed occupations or O*NET-specific extensions within a broader SOC category.

sional background, prior work experience, and demographic attributes—from the LinkedIn profiles, using the founder-specific LinkedIn URLs provided in Crunchbase.

Although narrower in coverage and skewed toward more visible or venture-oriented startups, the merged Crunchbase–LinkedIn database provides the most granular information on founders' backgrounds, occupational histories, and early team composition.

2.3 Measuring GenAI Exposure

The extent to which startups can utilize GenAI depends on the nature of the tasks performed in their industry. Our primary proxy utilizes the Anthropic Economic Index (AEI) to measure task reliance on GenAI and leverages the Bureau of Census O*NET database to aggregate tasks at the occupation and industry levels.

Specifically, the AEI is constructed through a privacy-preserving analysis of millions of real-world conversations on Claude.ai. It measures the percentage of total queries involving a specific task among all queries to Claude during a week, including API queries, which may indicate business usage, as well as individual queries. Such an index allows us to capture the extent to which a task can be performed by AI 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.

Prior studies typically infer task exposure from the similarity of patents' textual descriptions to actual job descriptions involving AI (Eloundou et al., 2024; Hampole et al., 2025), from expert classifications and crowdsourced assessments (Felten et al., 2019; Brynjolfsson et al., 2018; Shao et al., 2025), or by surveying actual workers (Humlum and Vestergaard, 2024). Our measure, based on the AEI, leverages actual model queries to provide a more data-driven characterization of AI exposure. For our empirical exercise, unlike measures that rely on current job descriptions, the AEI also has the advantage of being forward-looking, as it measures how technologically savvy individuals currently utilize GenAI. It thus provides insight into how industries and occupations may be affected as GenAI becomes more widespread, and is particularly suitable for gauging AI use in new business ventures.

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 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.

Since our other data sources provide different industry classifications or levels of granularity, once we obtain the exposure measure at the four-digit NAICS industry level, we further aggregate it in some of our tests. In particular, we aggregate at 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, and it does not use a standard industry classification. 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 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.

For these reasons, we adapt our GenAI exposure proxy and construct a quasi-industry measure of GenAI exposure based on the pre-GenAI occupational composition of pre-GenAI startups, as observed in Revelio Lab, within an industry category. Let $AEI_{j\tau}$ denote firm j'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):

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

This measure provides a quasi-industry proxy for pre-GenAI technological exposure, reflecting both firms' occupational composition and their category-group affiliations prior to the diffusion of GenAI. The measure reflects firms' technological task composition before the GenAI era and serves as a proxy for sector-level exposure to GenAI. Firms that belong to multiple category groups are assigned the simple average of the corresponding Exp_c^{pre} values. Thus, we obtain a measure of technological exposure to GenAI that accounts for firm-level heterogeneity from multi-category affiliations and reflects the average GenAI AEI intensity of the category groups to which a firm belongs.

2.4 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.

Figure E.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 exposure to GenAI differs significantly in the merged Crunchbase-Revelio sample. The firms in this sample are natural candidates for venture capital funding and are predominantly in the IT sector, as shown in Table A.4, which lists industry categories in Crunchbase. It is unsurprising that all of them have some tasks that can be performed through GenAI. Importantly, 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. All these examples illustrate that the startups in our sample appear to be AI-native, meaning they develop new technologies, and that AI is central to their business models.

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 7.1 that our results are robust to measure task exposure to GenAI using conversations with an alternative widely used AI platform.

2.5 Summary statistics

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 Infogroup to examine changes in industry dynamics. At the four-digit level, while almost 80% of industry-years experience entry, the number of firms entering across industries varies widely. We also consider the number of entrants with sizes below/above the industry median at entry in a given year. Smaller entrants are relatively more frequent in our data. We will ask whether this pattern is accentuated by an industry's exposure to GenAI as this technology becomes more widely available.

We also consider, for each industry and year, the number of geographical areas experiencing new firm formation, moving from states to combined statistical areas to counties. There is substantial variation in the number of states, combined statistical areas (CSAs), and counties experiencing entry across industries, as well as in the concentration of firm entry across these geographical areas, indicating that agglomeration economies vary across industries. Again, we will test how these patterns vary with exposure to GenAI as the technology becomes more widely diffused.

Panels C and D summarize key characteristics of the merged Crunchbase-LinkedIn sample, focusing on startups and their founders, respectively. Specifically, Panel B presents the main startup characteristics at the end of the founding year. Naturally, with fewer than five employees, startups are small, though there is significant variation. We also present the number of occupations that startup employees report on LinkedIn at the end of the founding year. Variation in the number of occupations, which is, by construction, smaller than the number of employees, may depend on industry characteristics and, as we will test, also on the diffusion of GenAI.

While Panel C summarizes the characteristics of all start-ups at the end of the funding year, it is important to note that Crunchbase underreports new startups that have not yet gained visibility or external funding. Furthermore, LinkedIn coverage may be incomplete if employees do not list newly founded or informal projects, especially at very early stages of development. This suggests that the post-2023 sample may be biased toward more established or better-performing startups. If anything, this bias should attenuate the estimated effect of GenAI on making startups leaner: firms that are already successful are unlikely to appear smaller or hire fewer employees than startups in the same industry before the diffusion of GenAI.

Panel D describes the founders' characteristics. We can identify employee information on LinkedIn for approximately one-third of the startups. Hence, there are fewer observations of startup founders. Founders are disproportionately male and white. In addition to examining how their demographic characteristics change with the industry's exposure to GenAI, we also use proxies for founders' prior occupational exposure to tasks that can be performed with GenAI after its diffusion to evaluate how this affects their propensity to start a business. Specifically, we consider the extent to which their previous occupations can be performed with GenAI on average, as well as the GenAI exposure of their most recent occupation or the occupation with the greatest exposure.

3 High-Frequency Evidence from Projected Businesses Formation

Using data from the BFS, we provide preliminary evidence on how business creation responds to the diffusion of GenAI. We test whether industries in which relatively more tasks are affected by GenAI experience a larger increase in entrepreneurial entry in the months following its diffusion. We thus exploit variation in industry-level exposure to GenAI, as captured by the aggregated AI exposure index, to compare changes in new business formation before and after the diffusion of GenAI technologies using a difference-in-differences specification.

We estimate the following empirical model:

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

where the outcome variable, Y_{nmt}, is the number of projected business formations in 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 release in November 2022—marking the onset of the modern GenAI era, characterized by the rapid diffusion of foundation models and the widespread integration of generative tools into product development.

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 industry and year-month level, to account for potential dependence across 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 formations, including: (i) All business applications (column 1); (ii) The 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); (iii) The corporations' business applications (column 3). The estimated effects are not only statistically significant but also economically meaningful: In industries with exposure to GenAI above the median, we observe a 3.8% monthly increase in business applications (column 1) and in business applications that are 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% monthly increase in industries with exposure to GenAI above the median after the diffusion of GenAI.

Supporting the idea that we observe transitional effects, we observe substantial positive effects of the diffusion of the new technology in industries with high GenAI exposure, also for *Projected Business Formations* (PBF), which are Census projections of business propensities to become employers within four to eight quarters based on historical transitions (columns 4 and 5). This evidence suggests that GenAI could significantly impact entrepreneurial activity as it becomes more widely adopted.

To address concerns that we may capture pre-existing trends, we estimate differences in projected business formation between industries with high and low exposure to GenAI over time. The dynamic patterns presented in Figure 1 show that the positive effects of GenAI diffusion begin to emerge in late 2022, shortly after the release of early foundation models such as *Stable Diffusion* and *ChatGPT*. Importantly, there is no evidence of systematic pre-trends prior to this point, supporting the validity of the difference-in-differences identification. The post-2022 increase is concentrated in industries with high pre-existing AI exposure, consistent with GenAI spurring entrepreneurial activity in technologically aligned sectors.

4 Firm Entry and Industry Structure

4.1 Baseline Results on Firm Entry

While the BFS data allow us to study changes in entrepreneurial activity and track new business formations at high frequency, they prevent us from using granular industry information and drawing conclusions about the effects of GenAI on the industrial structure. To achieve greater granularity in industry classification and obtain firms' locations and size information, we consider newly established, single-location firms from InfoGroup. We treat 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 and construct several measures of new firm formation. Specifically, we compute: (1) the total number of new firms, to assess whether generative AI has enabled greater startup formation; (2) the geographical distribution of new firms, to examine whether entry has become more spatially diverse; and (3) the number of new firms below a size threshold, to test whether newly founded firms have become smaller or leaner in the post-GenAI period.[7]

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 the following difference-in-differences specification:

$$Y_{nt} = \alpha + \beta\big(AEI_n^{\text{NAICS4}} \times Post_t\big) + \delta_n + \delta_t + \varepsilon_{nt}, \tag{4}$$

where Y_{nt} denotes the outcome for industry n in year t, measured as the number of newly established firms or the other outcome variables listed above; AEI_n^{NAICS4} the four-digit NAICS industry's GenAI exposure; and $Post_t$ is an indicator equal to one for the post-GenAI period (2023–2024) and zero otherwise. All specifications include industry (δ_n) and year (δ_t) fixed effects. Standard errors are clustered at the four-digit NAICS level.

We estimate Equation 4 using industry-level yearly data from 2020–2024. The coeffi-

[7] The number of newly established firms appears lower in 2024 relative to prior years, most likely reflecting incomplete data coverage for that year, as the dataset was downloaded in mid-2025. As noted earlier, the temporal selection of the sample is likely to bias our results downward.

cient β on the interaction term ($AEI_n^{\text{NAICS4}} \times Post_t$) captures how patterns of firm creation evolved after the diffusion of GenAI in more-exposed industries *relative* to less-exposed ones.

Table 3 reports the estimates. We compare industries with exposure to GenAI above/below the median, as in the earlier analyses. We also include specifications in which we enter the GenAI exposure interaction variable in quartiles, coded as a categorical variable that ranges from 0 (1st quartile) to 3 (4th quartile), so the coefficient of the interaction term ($AEI_n^{\text{NAICS4}} \times Post_t$) captures the change in the gap per quartile. We find that in more AI-exposed industries, the probability of observing any new firm 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 of Table 3). 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-related tasks prior to 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% in industries with GenAI exposure above the median following the introduction of GenAI. 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, leveraging 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 involving each task with the fraction of interactions occurring in a given mode to obtain automotive and augmenting interactions. Then we aggregate at the task and occupation level as we did for the construction of AEI_n^{NAICS4}. Figure 2 presents the estimates obtained from Equation 4 using industry measures of automotive and augmentative GenAI exposures. The estimates are qualitatively and quantitatively similar to the baseline, suggesting that both automotive and augmentative modes facilitate 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 has a positive effect on 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.

4.2 Geographic Dispersion of New Entrants

We next examine the geographic distribution of new entrants in industries more exposed to GenAI. Table 4 asks whether the number of geographic areas with entry in a given industry during a year increases post-GenAI and estimates Equation 4 to evaluate any changes. In addition to the number of geographic areas, we also consider the concentration of new entrants within each area. These variables can provide insight into how the use of GenAI is impacting the industrial structure.

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 CSAs (columns 2 and 5).

More states, CSAs, and counties record new firm formation in highly exposed industries after 2023 (columns 1 to 3), and the geographic Herfindahl–Hirschman Index (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 40% decrease in the county-level concentration of entry in industries more exposed to GenAI.

These patterns suggest a decline in geographic concentration and reduced entry barriers: technological advances, such as GenAI, may enable entrepreneurs outside traditional innovation hubs to enter AI-adjacent markets. In other words, GenAI appears to democratize access to new firm creation, both by lowering organizational scale requirements and by expanding participation across locations.

We also explore the mechanisms driving the increasing geographic dispersion of entrepreneurial activity. Agglomeration economies are often thought to be driven by local access to specialized labor and, more generally, local knowledge spillovers, as well as greater availability of funding due to proximity to venture capital firms (Ellison and Glaeser, 1997; Ewens and Farre-Mensa, 2022). GenAI may reduce agglomeration economies because technology can substitute for scarce, specialized human capital by automating certain tasks and augmenting the founders' capabilities, thereby reducing the need for local knowledge. Since startups exposed to GenAI also appear to be smaller, lower startup costs may reduce their funding needs and the benefits of locating in areas with historically concentrated VC investment.

Table 5 investigates whether both these channels are in operation. 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.[8] 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 specialization of local labor 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 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 refine this classification at the county level by aggregating the total dollar value of investments made by U.S.-based VC investors in each county in 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, the results—reported in Panel B of Table E.2 show stronger GenAI-induced entry in counties with limited VC activity.

4.3 The Size and Organization of New Entrants

Infogroup data enables us to examine the characteristics of new entrants. We present the results in Table 6. It shows that newly created firms in more exposed industries tend to be smaller. The increase in entry is concentrated among firms with initial employment

[8]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 E.2.

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 3 and 4). Using absolute size thresholds (e.g., 3/5/10 employees) produces similar results (Table E.3). This pattern suggests that GenAI diffusion is associated with a shift in composition toward smaller, leaner startups. One possible interpretation is that access to GenAI tools allows entrepreneurs to substitute technology for labor in the early stages of firm formation, reducing the need for large founding teams.

These findings are confirmed in the merged Crunchbase-Revelio sample, which allows us to examine not only the size of startups but also how the number of distinct roles they offer varies with the advent of Gen AI. To examine how startups' characteristics evolve in industries differentially exposed to GenAI, we estimate the following empirical model:

$$Y_{ft} = \alpha + \beta\big(Exp_c^{\text{pre}} \times Post_t\big) + \delta_t + \sum_{c \in \mathcal{C}} \phi_c D_{fc} + \varepsilon_{ft}, \tag{5}$$

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 (*Employees*), the number of unique occupations among the firm's employees by the end of the founding year (*Occupation count*), and the Herfindahl–Hirschman Index of occupational concentration within the firm, capturing within-firm task diversity (*Occupation HHI*).

The variable Exp_f^{pre} is the firm's ex-ante GenAI exposure, defined in Section 2.3, based on its category groups. As in the earlier tests, the indicator $Post_t$ takes the value once for the post-GenAI period ($t \in \{2023, 2024\}$) and zero for the pre-GenAI period ($t \in \{2020, 2021, 2022\}$). Thus, the coefficient β captures how the composition of a startup's workforce at founding 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.

All specifications include founding-year fixed effects, δ_t, and industry category dummies, D_{fc}, which take a value equal to one if firm f is associated with category group c. Because firms in Crunchbase can belong to multiple industry category groups (typically 2-5), several of these dummy variables may be 1 simultaneously. Since industry categories are not mutually exclusive, we use bootstrapping to estimate standard errors.

Table 7 shows that, relative to less-exposed peers, startups with higher ex-ante exposure (Exp_c^{pre}) founded in 2023–2024 exhibit *fewer employees*, *fewer distinct occupations*, *higher occupational concentration (HHI)* among their employees at the end of the founding year. The estimates are largely invariant to whether we include founding-year fixed effects. Importantly, the parameter estimates are not only statistically significant but also economically meaningful. The average startup in our sample has 4.6 employees. Thus, the parameter 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. These post-GenAI patterns are unlikely to be driven by shrinking sample coverage over time, because selection should reduce the probability of observing less visible startups. The most successful startups, which are more likely to be overrepresented towards the end of our sample, should have more employees, making it harder to detect the effects of GenAI.

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

Figure 3 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). It is evident that the frequency of startups with only one employee increases significantly in the post-period across 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 further confirms that sample selection is unlikely to drive our findings.

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.

5 Startups' Technological Characteristics and the Mechanisms of Technological Diffusion

The evidence we have presented so far is consistent with the hypothesis that GenAI reduces cognitive costs and diminishes skill complementarities, leading to smaller founding teams and greater occupational concentration. This suggests that GenAI favors automation and reduces entry costs. Mere automation, however, is unlikely to drive growth or significant, sustained productivity gains if startups do not introduce new products and technologies.

To shed light on this issue, we examine how GenAI is altering the technological characteristics of startups. GenAI can affect either the process or the products offered by startups. Specifically, 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 adopt leaner production processes that reduce dependence on specialized labor and external financing.[9] However, GenAI can also increase entry by facilitating the creation of new products and services and creating investment opportunities for entrepreneurs. If this were the case, we expect that at least some startups would be able to make AI the core of their operations and create new AI-powered products.

To gain insight into how GenAI is affecting startups' technological characteristics, we

[9]Open-source models such as Llama or DeepSeek provide a similar pathway.

utilize the business descriptions of more than 260,000 startups in Crunchbase. We develop two prompts to identify GenAI-enabled startups, defined as startups that use GenAI to automate their operations, and AI-native startups, defined as those that offer new products and services that rely on AI.

To analyze startups' business descriptions, we use GPT-5-mini as the primary model to execute our prompts. We also validate a subsample of the classification using GPT-5-nano and achieve model agreement of over 98%. The prompt to identify startups that are *AI natives* is designed to exclude firms that use AI merely as an add-on feature or as "API wrappers" that pass on the functionality of existing foundation models to users. Thus, AI-native startups are those that put a novel AI model at the core of their products and services. AI-enabled startups instead rely on AI to automate certain processes but offer products and services that do not rely on AI. We provide details on our prompts in Appendix B. While identifying AI-enabled startups is somewhat more challenging because not all firms disclose how they execute their processes, Panel B of Figure 4 shows the functions in which AI-enabled startups most frequently use GenAI applications and provides a validation for our prompt. AI-enabled startups appear to use GenAI for product management, data analytics, marketing, etc. All these functions appear to involve tasks with high exposure to GenAI.

Panel A of Figure 4 shows that the share of both AI-enabled and AI-native startups increases after 2022, with a steeper increase for AI-enabled startups, that is, startups that rely on GenAI to achieve leaner operations. However, the share of AI-native startups is significantly higher than that of AI-enabled startups, increases steadily over the sample period, and accelerates after 2022, consistent with a process of technological diffusion that has led to the emergence of new products and services offered by AI-native startups. Thus, while new entrants appear to use foundational models and automate operations at an increasing pace, the surge in entrepreneurial entry does not appear to be driven solely by reliance on ChatGPT and other models in order to automate and reduce entry costs. A significant proportion of startups is creating new products and services, which could be conducive to growth. The presence of a significant proportion of AI-native startups before the GenAI boom is because many of the startups we categorize as AI-native also use traditional AI technologies in their products. Reliance on GenAI models is likely to have reduced the fixed costs of creating and improving traditional AI models. This suggests that GenAI has increased the productivity of older AI technologies, exhibiting characteristics of general-purpose production technologies.

The differences in entry rates across industries with different exposure to GenAI appear to be driven by both GenAI-enabled and GenAI-native startups. In Figure 5, the share of GenAI-enabled startups in an industry is positively related to the industry's exposure to GenAI in 2023 and 2024, while the relation appears much flatter in 2020 and 2021, that is, before the release of mass foundation models and the dramatic improvements in the GenAI technology. The share of startups we identify as AI-native in each industry category group is also positively associated with that group's GenAI exposure. Figure 6 visualizes this relation for each of the years in our sample. We find a positive relationship in each sample year; importantly, the relationship becomes steeper over time, suggesting that technological diffusion may have led many startups to refine their AI models for market applications starting in the second half of 2022.

Overall, these findings suggest that while GenAI is certainly favoring entry by enabling startups to automate tasks, it is also enhancing their AI product offerings, with the potential to increase economic growth.

6 GenAI and Labor Demand of Startups

6.1 Aggregate Employment Effects in Entrepreneurial Firms

Our empirical 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 can mitigate concerns about widespread labor displacement and depressed labor demand, or instead accentuates them due to startups' smaller size.

To evaluate these aggregate effects, we need to explore whether job creation in startups that are relatively more exposed to GenAI has increased. To do so, we use QWI, a nationally representative administrative measure of employment and earnings for very young firms (age 0-1) 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 introduction of GenAI. The results in Table 8 show a clear and economically meaningful pattern: employment among age 0–1 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 technology adoption and cannot 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. To shed light on how GenAI is changing labor demand, we explore the occupational composition of startups and the characteristics and skills of their founding teams.

We investigate 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 by 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. Results are presented in Table 9. We find that 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 and reflect how startups in more-exposed categories reorganize their initial teams differently from those in less-exposed ones.

Our findings suggest a relative reallocation of employment in startups toward more GenAI-intensive tasks, rather than a uniform downsizing across all occupations. GenAI appears to reduce the minimum efficient scale of human labor at founding by (i) substituting for support and peripheral roles, (ii) compressing the skill bundle needed to reach a viable product, and (iii) enabling founders to externalize activities to tools, platforms, and short-term contractors. As a result, founding teams include individuals with a narrow set of GenAI capabilities for system integration, model deployment, or product implementation of generative tools. 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 employees' working hours in occupations using GenAI have increased following the diffusion of GenAI, presumably due to increased productivity (Jiang et al., 2025). In contrast, medium- and low-AEI occupations are less complementary to GenAI-related technologies, leading startups in more-exposed categories to scale back hiring in these roles.

Importantly, there are significant differences in the occupational composition of startups and mature companies. While high-AEI occupations appear complementary to 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). Our results in Table 9 suggest that the investment opportunities created by the new technology increase the demand for these roles in startups, mitigating concerns about widespread job displacement.

Figure 7 provides more granular evidence on this point. It 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 appear to have recorded the largest occupational gains, increasing 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 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 easily 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 8 provides a graphical representation of the occupations that experience the largest gains in employment at startups as a function of the occupation's exposure to GenAI and confirms this point. This is not only the case for the full sample of occupations (Panel a) but also for those in the top decile of the pre-period startup employment share (Panel b). Thus, demand for occupations with high exposure to GenAI in startups appears fundamentally different from that in large, mature companies, which benefit from coordination roles and require proportionally fewer employees to establish GenAI infrastructure. This neglected mechanism we highlight mitigates 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 high exposure to GenAI. To explore whether this is the case, for each founder k of firm f founded in year $t \in \{2020, \ldots, 2024\}$, we estimate:

$$Y_{kft} = \alpha + \beta\big(Exp_c^{\text{pre}} \times Post_t\big) + \delta_t + \sum_{c \in \mathcal{C}} \phi_c D_{fc} + \varepsilon_{kft}, \tag{6}$$

where the dependent variable, Y_{kft}, captures different characteristics of founder k of startup f in founding year t. Given our results on the jobs created by startups, we first examine 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 consider the GenAI exposure of the founder's last occupation prior to founding ($PriorAEI^{\text{last}}$); the average GenAI exposure across all the founder's pre-founding occupations ($PriorAEI^{\text{mean}}$); and the maximum GenAI exposure observed across all pre-founding occupations ($PriorAEI^{\text{max}}$). We also consider founders' educational attainment, using a categorical variable for the highest educational attainment, coded from 1 (high school) to 6 (Ph.D.).

Existing literature suggests that women and minorities are underrepresented among startup founders because they lack access to networks or face discrimination in the funding process, as well as other barriers (Ewens, 2023; Gornall and Strebulaev, 2025). We thus ask whether, by reducing dependence on cofounders' skills and external finance, GenAI favors the entry of historically underrepresented individuals into industries more exposed to GenAI. We thus consider the founder's gender and race, using the following indicator variables for demographic composition: *Female*, *White*, and *Asian*.[10]

As in the previous section, Exp_f^{pre} denotes the firm's GenAI exposure by industry category groups; we distinguish between occupations with GenAI exposure above and below the median. $Post_t$ is an indicator equal to one for the post-GenAI period ($t \in \{2023, 2024\}$) and zero for the pre-GenAI period ($t \in \{2020, 2021, 2022\}$). We include founding-year fixed effects (δ_t), and industry-category dummies in all specifications.

[10] Revelio estimates race based on individuals' names. We recognize that the classification can be noisy and that the noise may be particularly pronounced for minority groups. For this reason, we do not use data on Hispanics and Black people.

We estimate Equation 6 using founder-level data from 2020–2024. The coefficients on the interaction term ($Exp_f^{\text{pre}} \times Post_t$) capture how founder characteristics have changed after the diffusion of GenAI in startups belonging to more-exposed categories relative to less-exposed ones. The results thus describe differential shifts in founder composition across startups with varying levels of exposure to GenAI, rather than absolute changes over time.

Table 10 shows that founders of startups in more GenAI-exposed categories have higher values of exposure to GenAI in their past occupations. This not only reflects the founder's tendency to start businesses in industries where they already have experience, but 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.

This result is consistent with two non-mutually exclusive forces. On the one hand, individuals with prior experience in AI-intensive occupations may be better equipped to recognize and capitalize on opportunities created by GenAI technologies. On the other hand, 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 to adapt. The observed increase in founders with high-GenAI-exposure backgrounds, therefore, likely reflects both opportunity-driven entry and reallocation following task-level displacement.

In either case, the founders of businesses exposed to GenAI appear to leverage experience from previous occupations where many tasks can now be automated or augmented by GenAI, to capitalize on the new technology in their ventures. 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—capturing hard skills accumulated through past employment—Figure 9 provides complementary evidence on how the content of founders' self-described characteristics shifts following the diffusion of GenAI. Using an LLM-assisted pipeline described in Appendix D, 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.

These patterns suggest that, following the diffusion of GenAI, founders increasingly emphasize roles and capabilities focused on building, deploying, and scaling GenAI-enabled products, rather than on activities less central to GenAI-augmented production. Importantly, the close correspondence between the observed-only and observed-plus-inferred results indicates that these shifts are not driven solely by changes in narrative style but reflect substantive differences in the roles and capabilities founders highlight at entry.

Table 11 shows how founders' educational background and demographics evolve differentially in industries with GenAI exposure above the median. The traditional entrepreneur has typically been white and male, even though survey evidence indicates that since 1996, the share of new entrepreneurs who are white has declined and the share who are Asian, Black, and Latino has increased (Desai and Fairlie, 2021). The share of female entrepreneurs has been roughly stable at 40%. We control for these trends in entrepreneurship to increase racial diversity among entrepreneurs by including founding-year fixed effects.

The estimates in Table 11 show that the relative decline in white founders is more pronounced in industries more exposed to GenAI, suggesting a shift toward greater demographic diversity among entrants. However, the proportion of Asian founders, which is close to or slightly above the Asian share of the U.S. population, has also increased. We observe no difference in the proportion of female founders. At the same time, founders in industries more exposed to GenAI exhibit lower educational attainment after 2022, as indicated by a negative coefficient on the interaction term between the startup exposure to GenAI and $Post_t$ in columns 4 and 6, where the dependent variable is the founder's *Education*. Taken together, these results suggest a compositional shift in the founder pool following the diffusion of GenAI towards individuals with fewer formal credentials.

Overall, these results suggest that GenAI-related technologies may have reduced financial, technical, and operational barriers to starting new ventures, enabling a broader set of individuals to enter entrepreneurship in these areas.

7 Robustness

7.1 Alternative Measures of Exposure to GenAI

As discussed in Section 2.3, our proxies for industry and occupation exposure to GenAI avoid relying on current job descriptions or individuals' opinions about occupations, which may be heavily influenced by current practices. The distribution of queries by a selected sample of technologically savvy individuals can inform us about the future use of GenAI. Furthermore, it enables us to measure how technologically savvy entrepreneurs are leveraging the new technology in ways that are more relevant to startups.

However, the AEI relies on a specific platform, Claude.ai. One may thus wonder to what extent our results are informative on the broader use of AI through startups' own models and other platforms. 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 over 20,000 O*NET task statements, Microsoft's index covers O*NET 330 intermediate work activities (IWAs) and aggregates upward to occupations using the occupational activity weights. Each occupation's score is computed as the weighted average of these activity-level values, with weights corresponding to the importance of each activity in the occupation's task profile (based on O*NET intermediate work activities).

At the O*NET occupation level (approximately 750 occupations), the correlation between the two raw measures is about 0.24, indicating that the two indices capture different dimensions of AI exposure. When comparing their rankings, however, the correlation increases to 0.65, indicating that occupations with high exposure scores on one measure generally rank high on the other as well. This pattern suggests that, despite differences in construction, 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 correlation increases to 0.42, and the rank correlation approaches 0.7, suggesting strong consistency in relative exposure patterns across industries. Figure E.2 provides graphical evidence on the comparability of the distribution of the two measures.

Importantly, 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 10 and 11 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.

We use the AEI as our primary measure because it encompasses all 20,000 O*NET tasks and is based on a substantially larger sample of user interactions (over several million conversations, compared with roughly 200,000 Copilot interactions in Microsoft's dataset). In addition, the AEI enables us to distinguish between first-party and API usage, as well as between automation and augmentation contexts.

7.2 Types of GenAI Interactions

Figures 10 and 11 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 based on the September 2025 version of the AEI, which is based on conversations from August 2025, 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. As a result, prompts related to non-commercial activities—such as trip planning or entertainment—lack clear counterparts in ONET and therefore cannot be matched to occupational tasks. In contrast, the prompts we successfully map correspond to business-oriented or professional activities, reflecting that ONET's task taxonomy is inherently work-related. Consequently, the resulting 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 11 shows that the estimated coefficient is indeed slightly higher. Third, estimates are invariant if we do not weigh tasks for their importance. 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 a GenAI effect.

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 10 and Figure 11), 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 by creating new investment opportunities for AI-native startups and lowering entry barriers for AI-enabled startups. The occupational composition of startups in industries exposed to GenAI is also changing. Following the diffusion of the new technology, startups in GenAI-exposed industries appear to be smaller and to have a higher concentration of employment in occupations that can be performed by GenAI.

Our results show that GenAI has a profoundly different impact on the occupational composition of startups and established companies. While previous literature highlights a decrease in jobs in occupations highly exposed to GenAI at mature companies, startups appear to have a higher demand for individuals with prior experience in such occupations.

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.

Furthermore, if GenAI reduces entry barriers and facilitates the formation of lean firms, policymakers aiming to leverage the new technology to spur growth and employment could focus on removing remaining frictions to startup activity. This includes improving access to digital infrastructure and reducing compliance burdens and legal uncertainties for small businesses enabled by GenAI. Policies that support experimentation with lean startup models (e.g., regulatory sandboxes) may be especially valuable.

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.

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.

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, January). General purpose technologies 'engines of growth'? *Journal of Econometrics 65*(1), 83–108.

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

Brynjolfsson, E., D. Li, and L. Raymond (2025, May). 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? In *AEA papers and proceedings*, Volume 108, pp. 43–47. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

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

De Loecker, J., J. Eeckhout, and G. Unger (2020, 05). 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, May). 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.

Desai, S. and R. Fairlie (2021). Who is the entrepreneur? the changing diversity of new entrepreneurs in the united states, 1996-2020. *Ewing Marion Kauffman Foundation, April 13*.

Dinlersoz, E., T. Dunne, J. Haltiwanger, and V. Penciakova (2021). Business formation: A tale of two recessions. In *AEA Papers and Proceedings*, Volume 111, pp. 253–257. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

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. (2023). Chapter 6 - gender and race in entrepreneurial finance. In B. E. Eckbo, G. M. Phillips, and M. Sorensen (Eds.), *Private Equity and Entrepreneurial Finance*, Volume 1 of *Handbook of the Economics of Corporate Finance*, pp. 239–296. North-Holland.

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. In *Academy of Management Proceedings*, Volume 2019, pp. 15784. Academy of Management Briarcliff Manor, NY 10510.

Felten, E. W., M. Raj, and R. Seamans (2023, April). Occupational heterogeneity in exposure to generative AI. Ssrn working paper, SSRN. Date Written: April 10, 2023; Posted: April 19, 2023.

Gornall, W. and I. A. Strebulaev (2025). Gender, race, and entrepreneurship: A randomized field experiment on venture capitalists and angels. *Management Science 71*(6), 5308–5327.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013, May). 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.

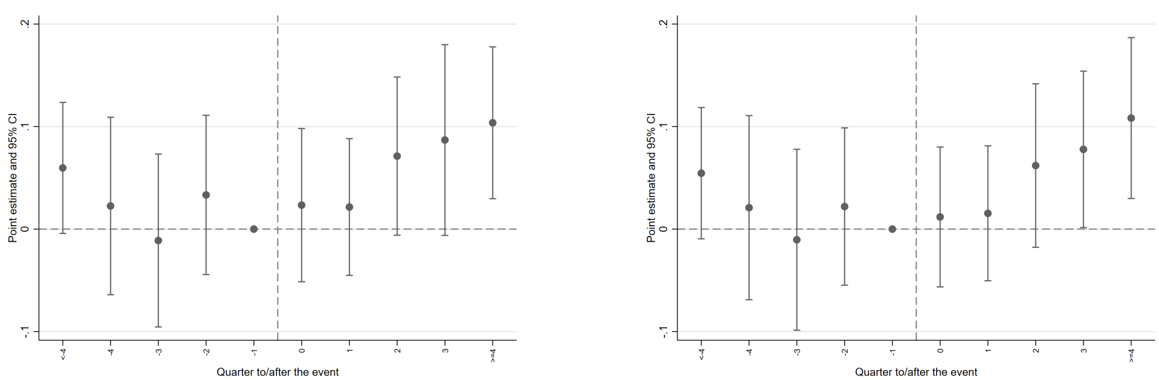
Jones, C. I. (2024, December). The AI dilemma: Growth versus existential risk. *American Economic Review: Insights 6*(4), 575–90.

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

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.

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*.

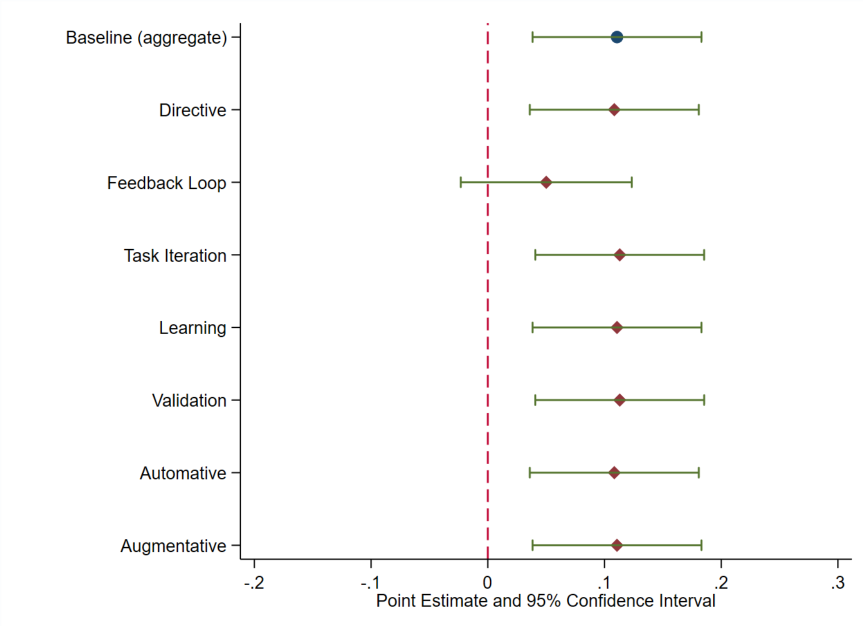
Figure 1: Dynamic Effects of GenAI Introduction on Business Formation



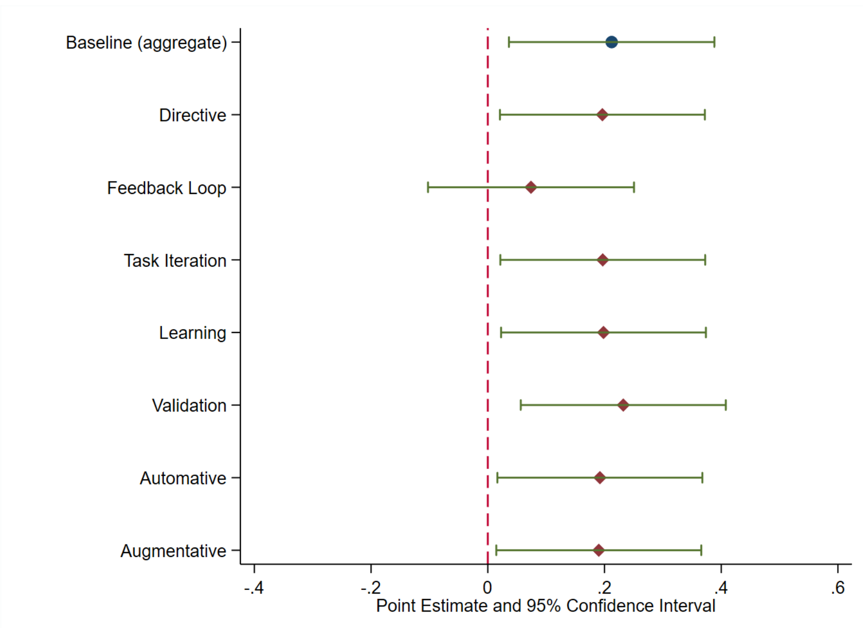
(a) Within 4 Quarters (PBF4Q) **(b) Within 8 Quarters** (PBF8Q)

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



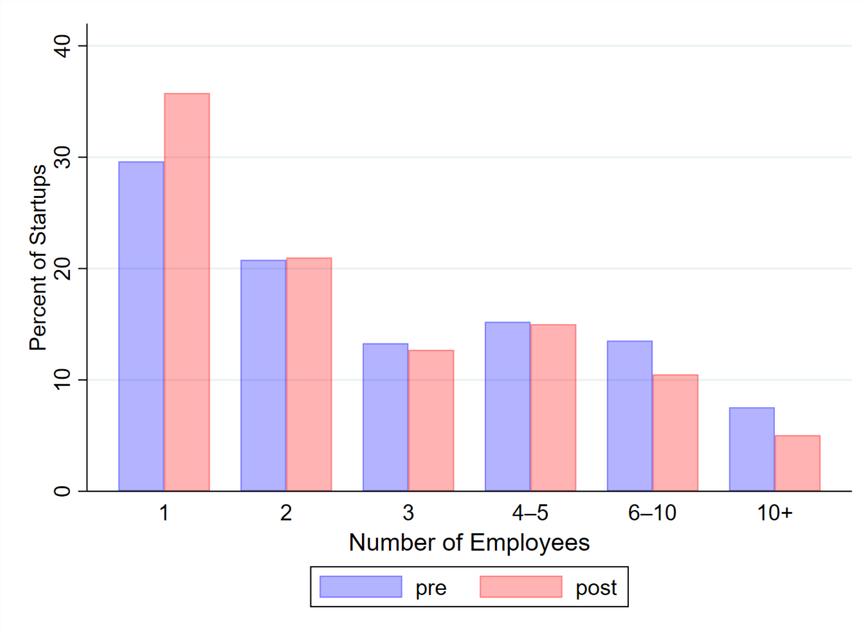
(a) Probability of Firm Entry



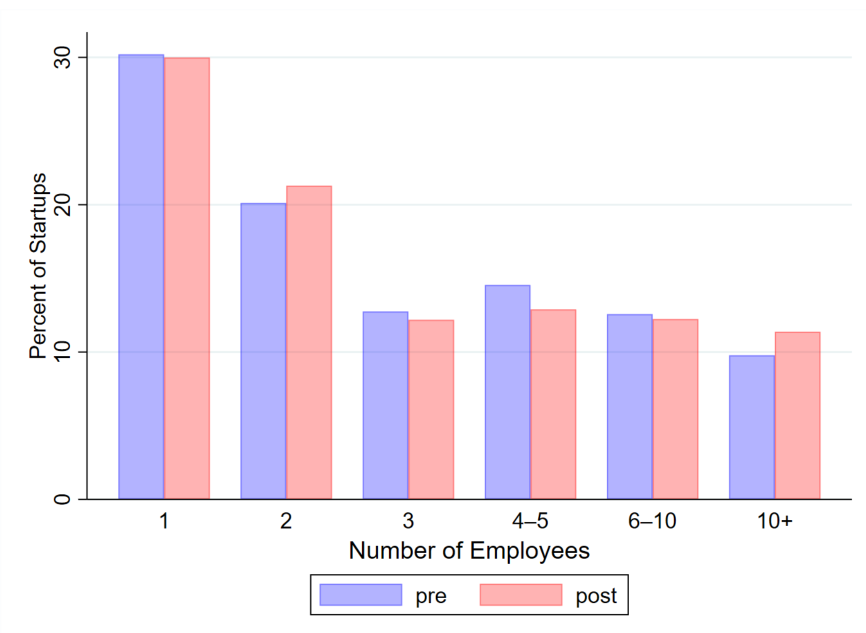
(b) # Firm Entry (log)

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: GenAI Introduction and the Changing Team Size Distribution of Startups



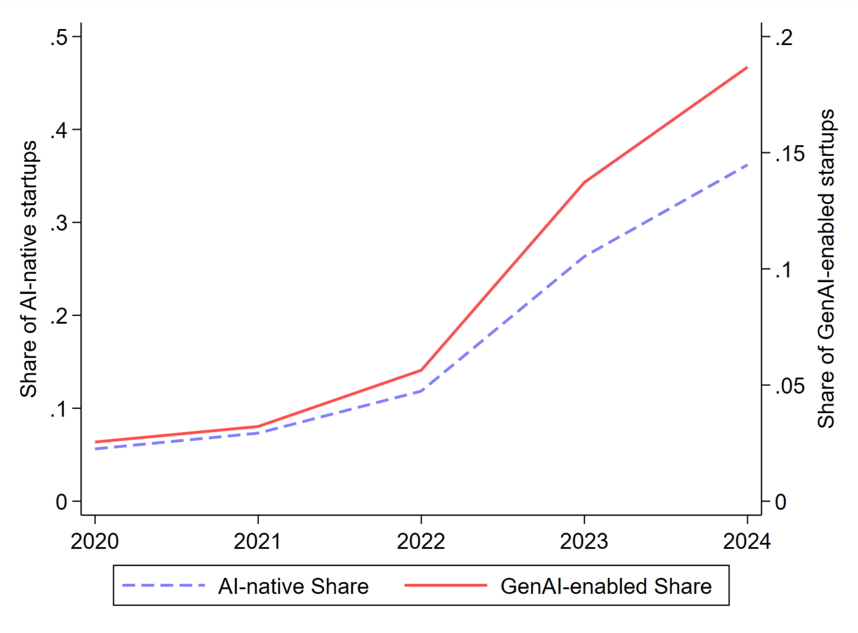
(a) High GenAI Exposure



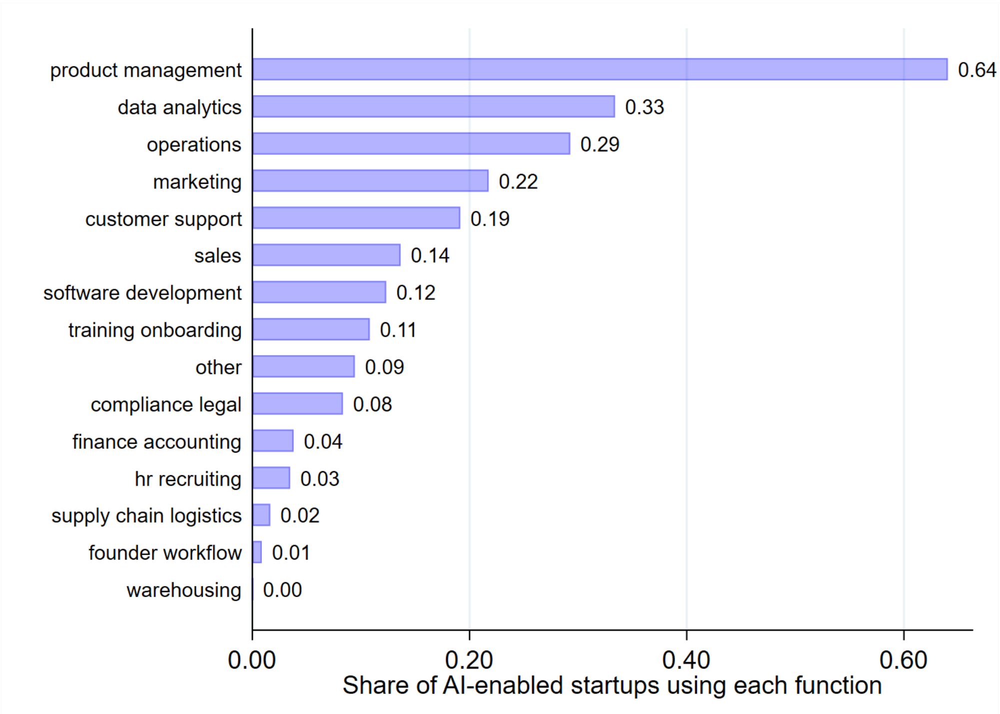
(b) Low GenAI Exposure

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 4: AI-Native and GenAI-Enabled Startups



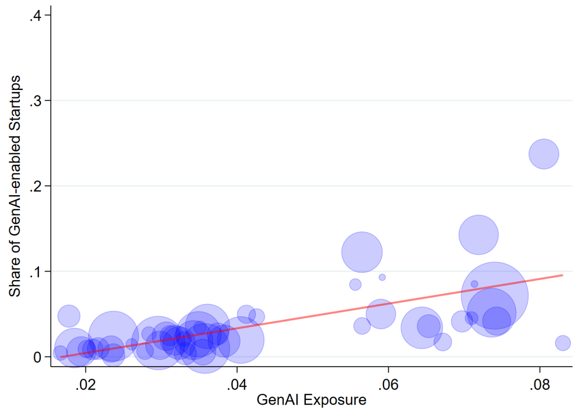
(a) Share of AI-Native and GenAI-Enabled Startups



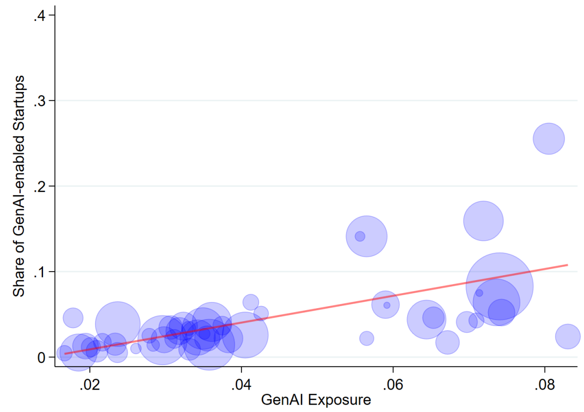
(b) Most Frequent GenAI-enabled Functions

This figure summarizes the evolution of AI-native and GenAI-enabled startups from 2020 to 2024. Panel (a) plots the share of startups classified as AI-native and GenAI-enabled each year. Panel (b) displays the most common GenAI-enabled functions in GenAI-enabled startups.

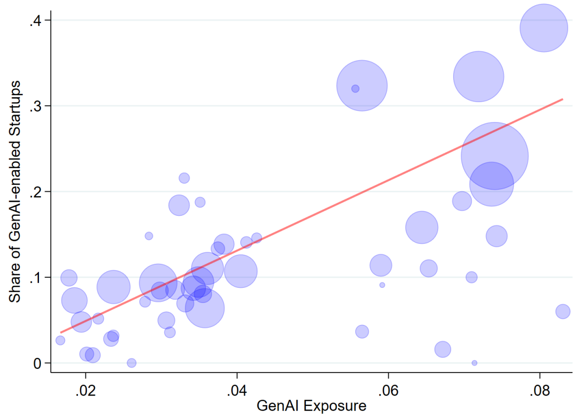
Figure 5: GenAI-Enabled Startups and Exposure to GenAI



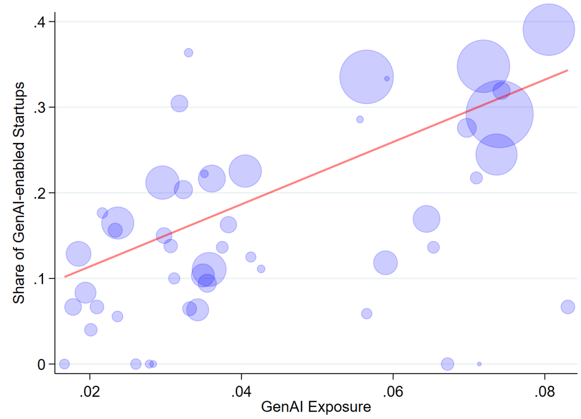
(a) 2020



(b) 2021



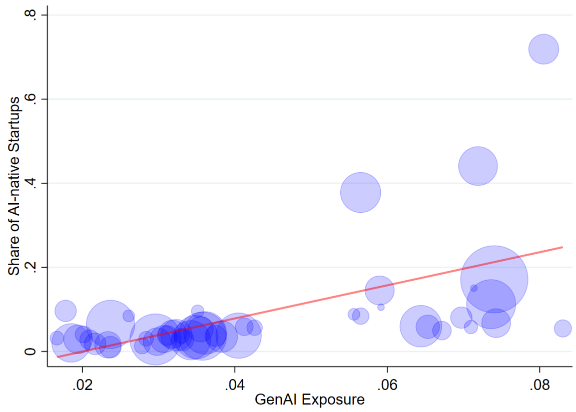
(c) 2023



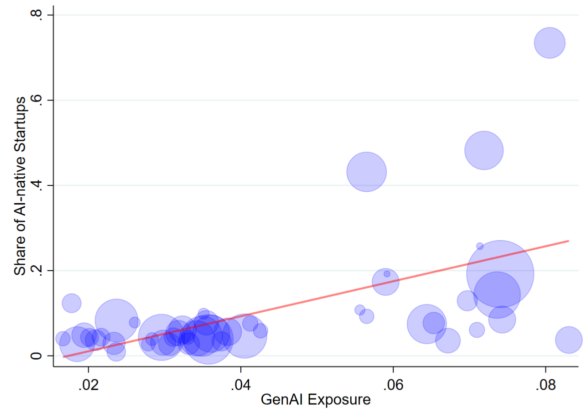
(d) 2024

This figure plots the share of GenAI-enabled startups 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.

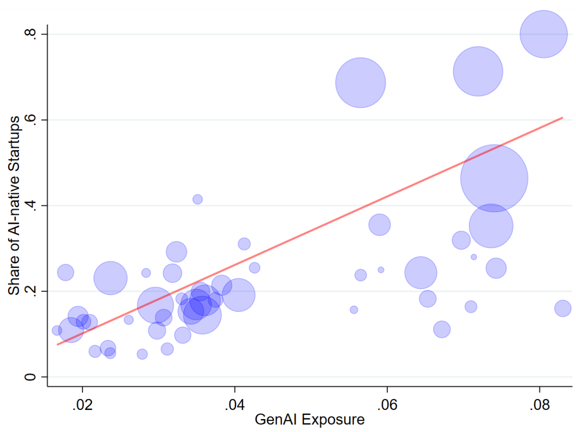
Figure 6: AI-Native Startups and Exposure to GenAI



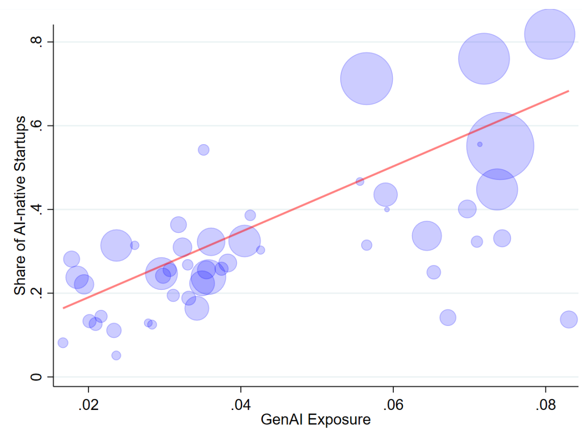
(a) 2020



(b) 2021



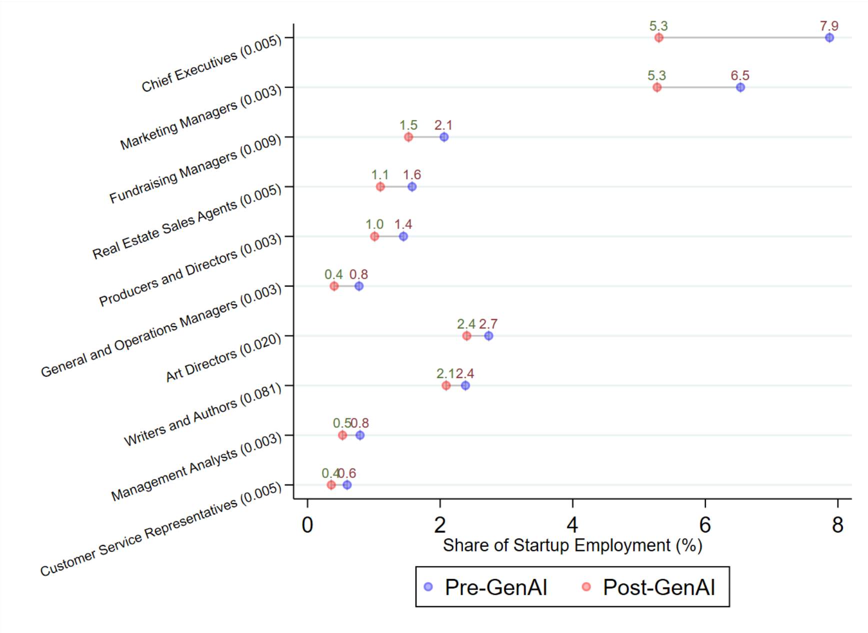
(c) 2023



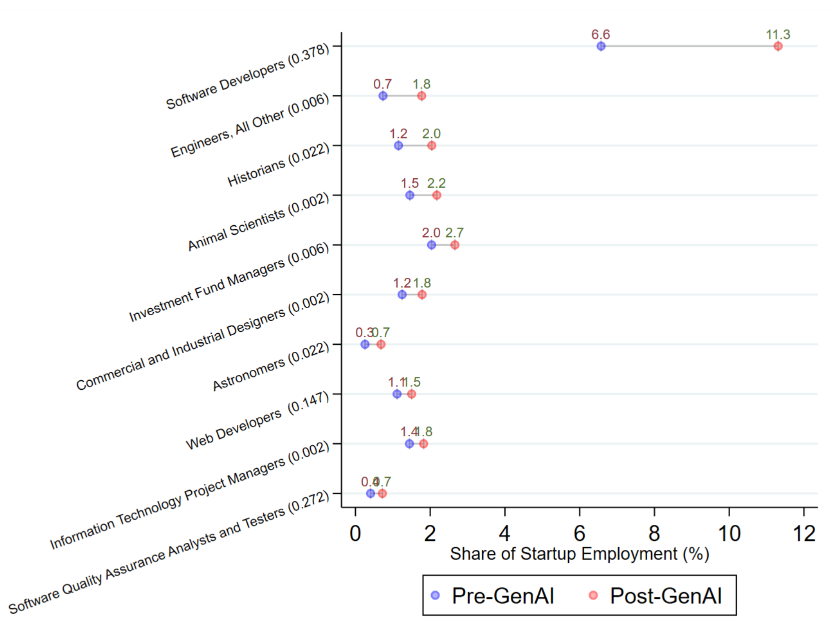
(d) 2024

This figure plots the share of AI-native startups 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 7: GenAI Introduction and Changing Startup Occupations



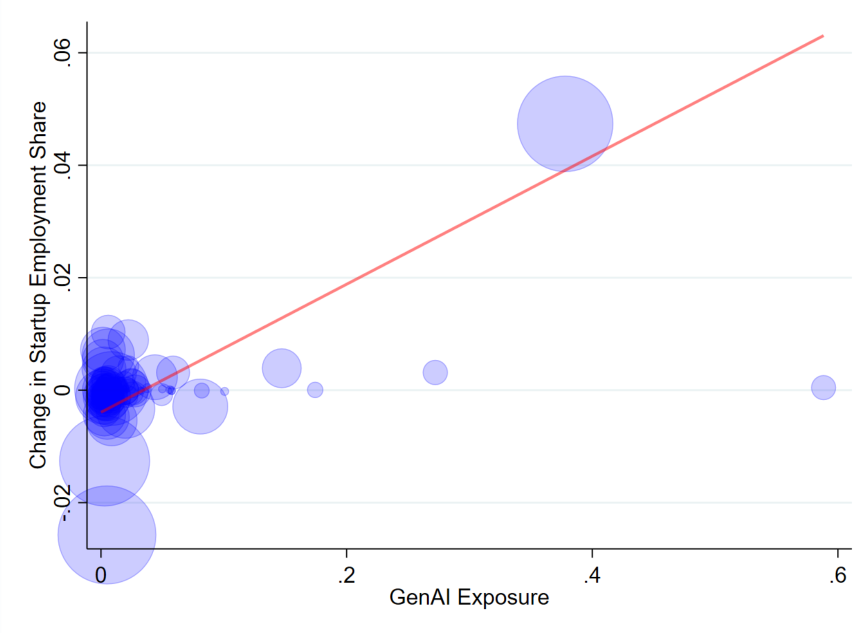
(a) Losing Occupations



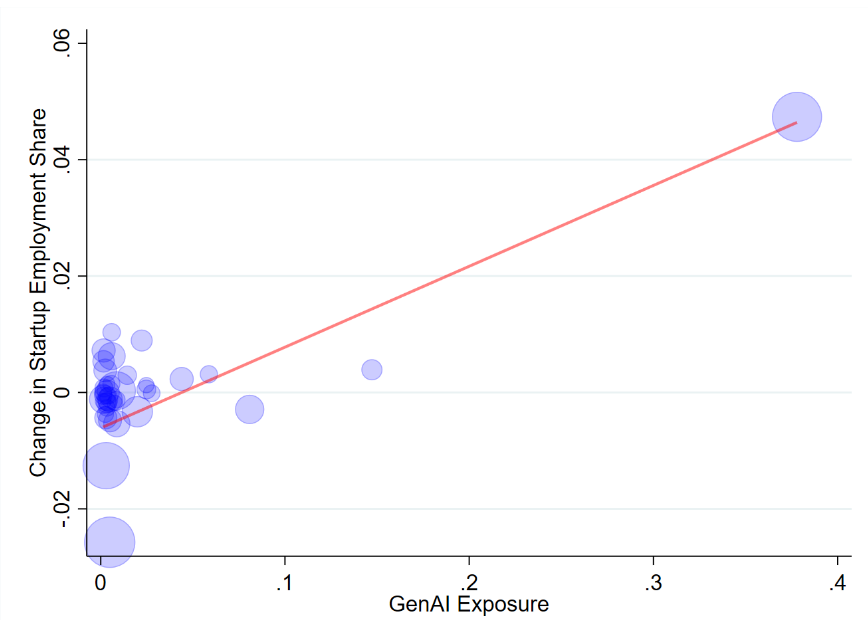
(b) Gaining Occupations

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 8: Changes in Startup Occupations vs. GenAI Exposure



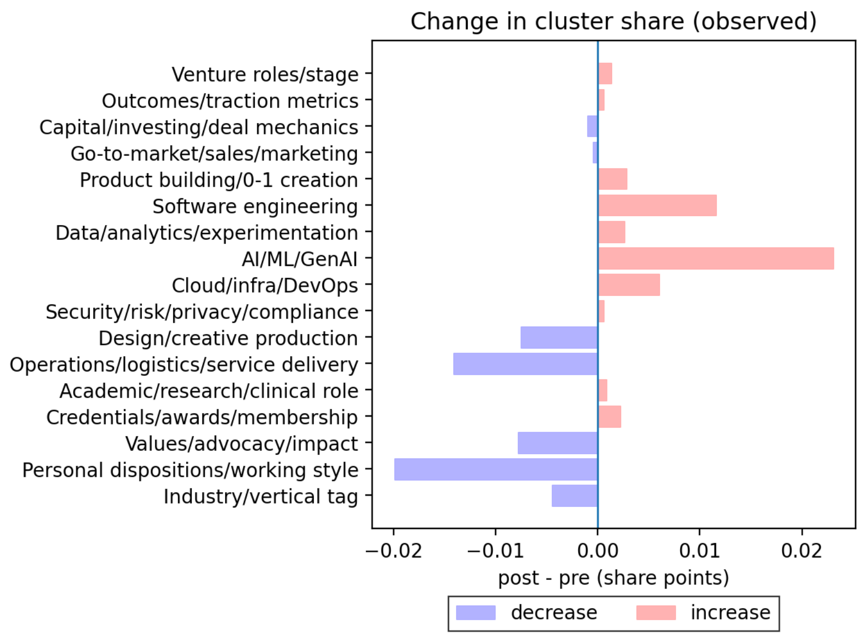
(a) All Occupations



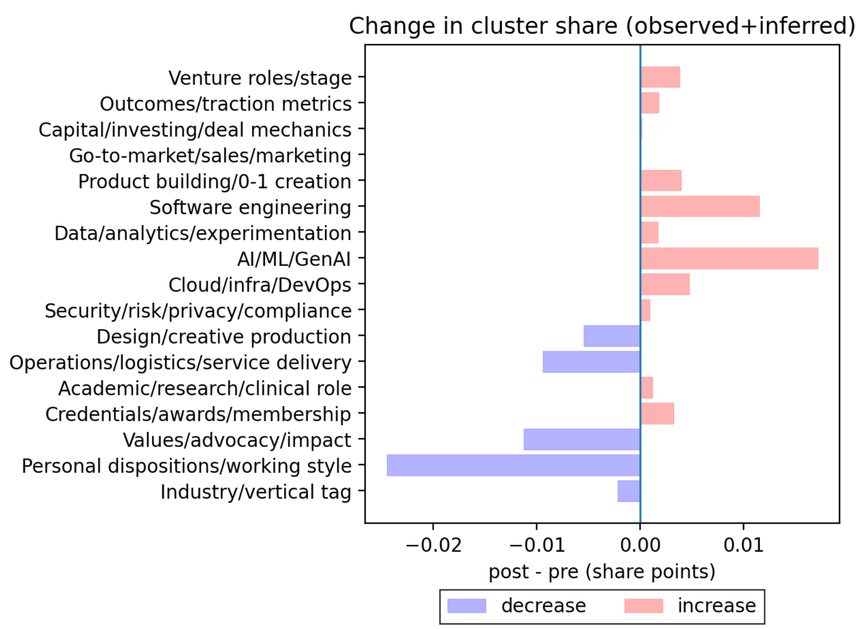
(b) Important Occupations (top decile)

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 9: GenAI Introduction and Founders' Changing Traits



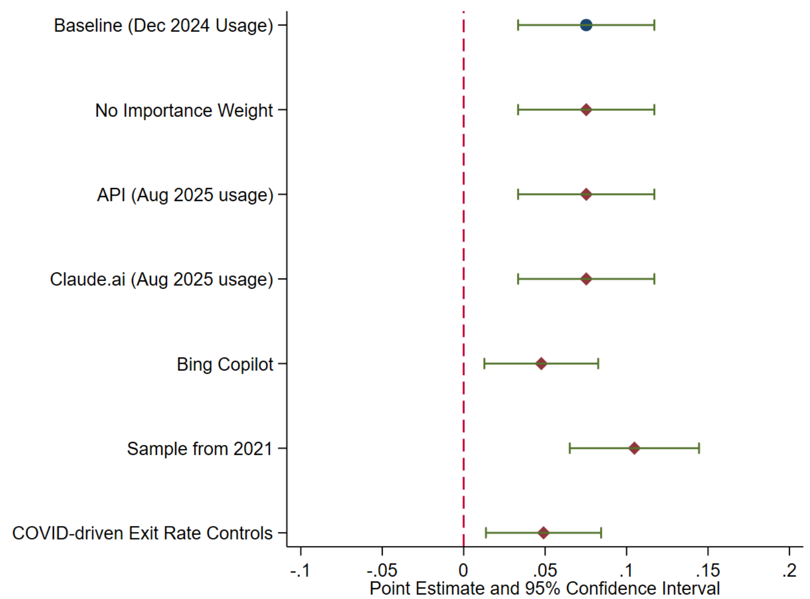
(a) Observed Traits



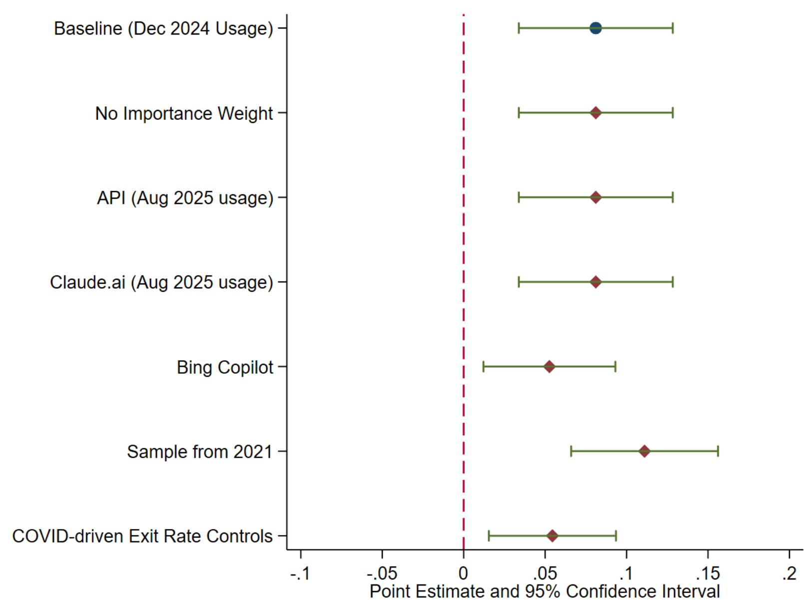
(b) Observed + Inferred Traits

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 D 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 10: Effects of GenAI Introduction on Business Formation—Robustness



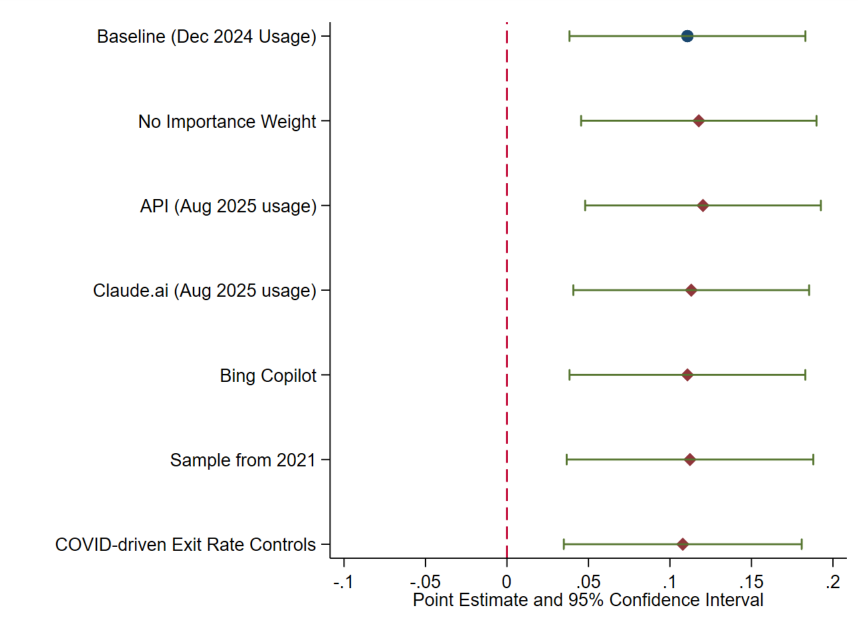
(a) Within 4 Quarters (PBF4Q)



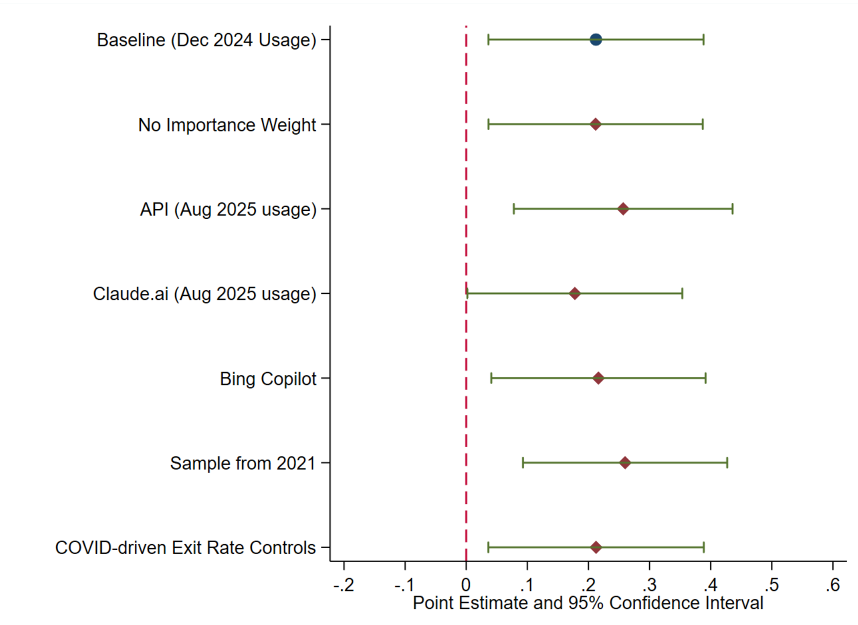
(b) Within 8 Quarters (PBF8Q)

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 11: 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 (Crunchabse-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
Male	14,892	0.778	0.373	0.995	0.000	0.699	0.997	1.000
White	14,892	0.635	0.383	0.829	0.000	0.227	0.972	1.000
API	14,892	0.162	0.321	0.004	0.000	0.001	0.075	0.999
Education (1-6)	11,811	3.915	1.174	3.000	1.000	3.000	5.000	6.000
AEI (most recent)	12,929	0.047	0.112	0.005	0.000	0.003	0.019	0.378
AEI (mean)	12,929	0.048	0.095	0.008	0.000	0.004	0.023	0.378
AEI (max)	12,929	0.092	0.157	0.020	0.000	0.006	0.062	0.588
AEI	14,892	0.047	0.016	0.046	0.019	0.033	0.061	0.074

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 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 year fixed effects are included in all specifications. Standard errors are obtained via a bootstrap procedure with 500 repetitions, resampling at the sector and year-month 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	1430	1430	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	1430	1430	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	1430	1430	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)–(2) examine firms below the median size threshold, and Columns (3)–(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]
Category Group Controls	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	rest	0-1	rest
AEI (above median) × Post_2023	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 quarterly 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***	0.022	-0.241***	-0.760***
	[0.013]	[0.018]	[0.064]	[0.040]	[0.053]	[0.204]
Category Group Controls	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	0.023***	0.016***	0.033***	0.019***	0.012**	0.023***
	[0.005]	[0.004]	[0.007]	[0.005]	[0.005]	[0.007]
Post_2022	-0.002	-0.002	-0.005			
	[0.003]	[0.002]	[0.004]			
AEI (above median)	0.046***	0.047***	0.086***	0.002	0.003	0.006
	[0.002]	[0.002]	[0.003]	[0.004]	[0.003]	[0.005]
Category Group Controls	NO	NO	NO	YES	YES	YES
Year FE	NO	NO	NO	YES	YES	YES
Obs	12,929	12,929	12,929	12,929	12,929	12,929
Adj R2	0.049	0.067	0.083	0.072	0.101	0.126

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 Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.	Male	White	Asian	Education	Male	White	Asian	Education
AEI (above median) × Post_2022	-0.026	-0.060***	0.069***	-0.155**	-0.018	-0.035*	0.039**	-0.196***
	[0.021]	[0.023]	[0.019]	[0.073]	[0.021]	[0.021]	[0.018]	[0.070]
Post_2022	0.024	0.012	0.000	0.166***				
	[0.018]	[0.018]	[0.012]	[0.063]				
AEI (above median)	0.086***	-0.078***	0.082***	-0.128***	0.041***	-0.01	0.013	-0.072*
	[0.006]	[0.007]	[0.006]	[0.023]	[0.012]	[0.012]	[0.010]	[0.039]
Category Group Controls	NO	NO	NO	NO	YES	YES	YES	YES
Year FE	NO	NO	NO	NO	YES	YES	YES	YES
Obs	14,892	14,892	14,892	11,811	14,892	14,892	14,892	11,811
Adj R2	0.012	0.013	0.023	0.004	0.045	0.023	0.043	0.116

This table reports the estimated effects of GenAI introduction on the demographic and educational composition of startup founders, 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. Columns (1)–(3) report indicators for whether the founder is male, White, or Asian/Pacific Islander (Asian), respectively, while Column (4) reports educational attainment, coded from 1 (high school) to 6 (doctorate). *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)–(4) report estimates without category-group controls or year fixed effects, while Columns (5)–(8) 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.

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

A Examples of Exposed Tasks/Occupations/Industries/Startups **A-2**

B Identifying AI-Native Startups **A-7**

C Identifying GenAI-Enabled Startups **A-8**

D LLM-Assisted Founder Characteristics Comparison **A-10**

E Additional Tables and Figures **A-14**

List of Appendix Figures

E.1 GenAI Exposure Distribution at the Industry and Startup Levels . . . A-17
E.2 Exposure Based on Microsoft Bing Copilot . . . A-18

List of Appendix Tables

A.1 Examples of O*NET Tasks by Share of Claude.ai Conversations . . . A-2
A.2 Examples of High and Low GenAI Exposure Occupations (O*NET Level) . . . A-3
A.3 Examples of High and Low GenAI Exposure Industries (4-Digit NAICS Level) . . . A-4
A.4 GenAI Exposure by Crunchbase Category Group . . . A-5
A.5 Top Ranked Startups based on GenAI Exposoure . . . A-6
D.1 Cluster description and representative examples (1/2) . . . A-12
D.2 Cluster description and representative examples (2/2) . . . A-13
E.1 Summary Statistics (QWI) . . . A-14
E.2 GenAI Introduction and the Geographic Distribution of Entry: Channels *Alternative Measures of Labor Specialization and VC hubs* . . . A-15
E.3 Effect of GenAI Introduction on the Size Composition of New Firms *Absolute Size Thresholds* A-16

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 Exposoure

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-Native Startups

We provide a prompt to large language models to classify each startup as `AI_native = 1` (AI-native) or `AI_native = 0` (non–AI-native). The prompt instructed the model to determine whether a startup's core product or service fundamentally depends on artificial intelligence or machine learning, and to produce a structured output line containing not only the binary classification but also additional metadata describing the reasoning behind the decision. Specifically, for each startup the model was asked to:

- assign a discrete confidence score (`Confidence_1to5`) from 1 (lowest) to 5 (highest),
- provide exactly three concise reasons justifying the classification (`Reasons_3_points`),
- indicate which input fields were used in forming the judgment (`Sources_used`),
- and include a short self-evaluation statement verifying or questioning its own decision (`Verification_critique`).

The prompt includes also few-shot examples to guide the reasoning. Note also that classification explicitly treats AI wrappers and API integrations around existing foundation models as *non–AI-native*, ensuring that the measure captures startups whose core products would not exist without AI or machine learning.

We applied this prompt to a large-scale dataset of more than 260,000 startup profiles obtained from *Crunchbase*. The sample is restricted to U.S.-based startups with non-missing founding dates in 2015 or later. The classification procedure was executed using *GPT-5-mini* as the primary model, chosen for its balance between accuracy and computational efficiency. To assess the robustness and replicability of the classifications, we additionally ran the same prompt on a representative subset of the data using an independent model, *GPT-5-nano*. The inter-model agreement rate between the two models was 98.36%, indicating a very high level of consistency in identifying AI-native versus non–AI-native startups.

C Identifying GenAI-Enabled Startups

We classify each startup as `GenAI_enabled = 1` (GenAI-enabled) or `GenAI_enabled = 0` (non–GenAI-enabled) using a structured prompt administered to a large language model. The objective is to determine whether a startup uses GenAI in any of its business functions—not whether it builds GenAI models themselves. The classification focuses exclusively on adoption of GenAI capabilities, such as text, code, image, audio, or video generation, conversational LLM-based assistants, or other forms of generative content creation, based on the startup's Crunchbase profile.

The prompt evaluates each startup using a hierarchical evidence framework: (i) the long business description (when available), (ii) the short description, and (iii) category tags as auxiliary signals. Importantly, the classifier is instructed to base all judgments solely on the text provided in the Crunchbase profile, without relying on external knowledge or industry stereotypes. The procedure explicitly excludes pure GenAI toolmakers whose entire product consists of providing an LLM or generative model, unless the profile also describes internal uses of GenAI in other functional areas.

For each startup, the model produces a structured, machine-readable output containing multiple layers of assessment:

- three parallel adoption decisions—*strict*, *moderate*, and *lenient*—each corresponding to well-defined evidentiary thresholds. For example, the *strict* criterion requires explicit mention of generative capabilities or unambiguous descriptions of generative tasks;
- a function-level breakdown mapping any detected GenAI use to standardized business functions (e.g., `marketing`, `customer_support`, `software_development`, etc.);
- and a concise reasoning field documenting which textual components supported the decision.

The strict mode is designed to avoid false positives by limiting GenAI adoption to explicit references such as "GPT-based chatbot," "LLM-powered assistant," "automatically generates marketing copy," or other phrases indicating generative content production. Moderate and lenient modes allow incremental inference when the wording strongly implies — but does not explicitly state — generative capabilities.

We apply this method to the same large dataset of U.S.-based startups used in our AI-native classification exercise. Classification is performed using *GPT-5-mini*.

D 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
LLM extracts founder traits from LinkedIn text

↓

Step 2: Clustering
LLM induces bottom-up clusters from keywords

↓

Step 3: Trait-to-Cluster Mapping
LLM assigns each trait to clusters using taxonomy rules

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 need to provide the text quote based on which it extrated 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 D.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 D.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	

E Additional Tables and Figures

Table E.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 E.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	1430	1430	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	1430	1430	1,430	1,430
Adj R2	0.491	0.58	0.89	0.81

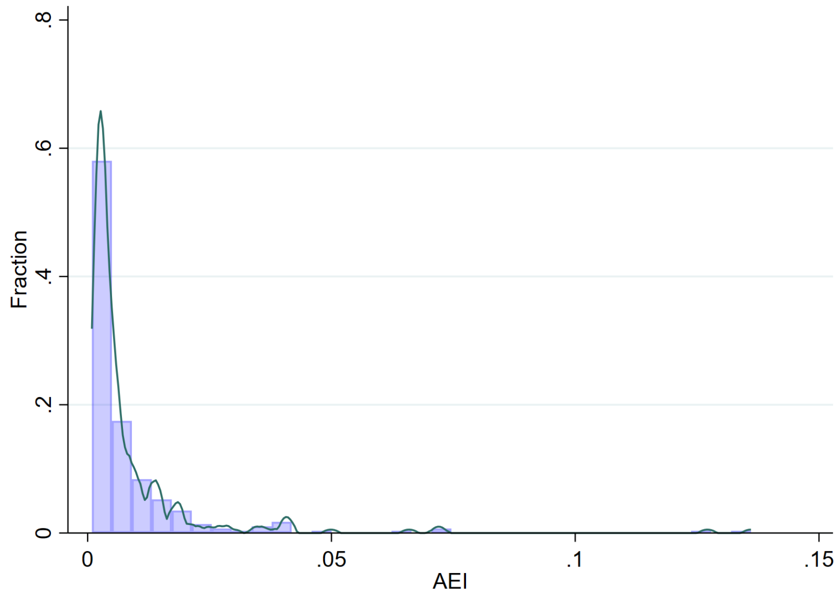
This table examines two channels through which the introduction of GenAI may have contributed to more geographically dispersed firm entry with alternative measures of labor specialization and VC hubs, using establishment-level data aggregated from the InfoGroup firm panel. The unit of observation is 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 E.3: Effect of GenAI Introduction on the Size Composition of New Firms
Absolute Size Thresholds

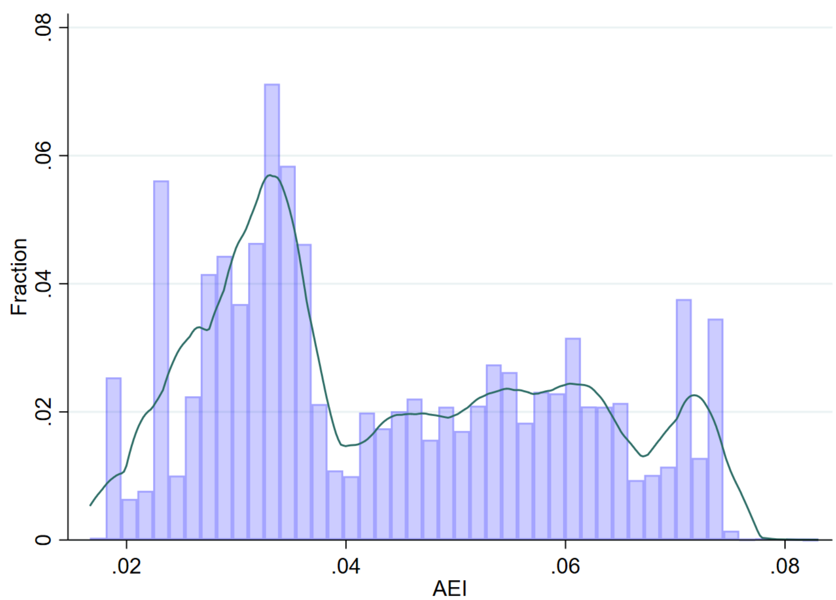
	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var.	# firm entry (log)					
	≤ 3 employees	> 3 employees	≤ 5 employees	> 5 employees	≤ 10 employees	> 10 employees
AEI (above median) × Post_2022	0.183**	0.031	0.174**	0.029	0.180**	0.06
	[0.075]	[0.087]	[0.079]	[0.085]	[0.082]	[0.083]
Industry FE	YES	YES	YES	YES	YES	YES
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 E.1: GenAI Exposure Distribution at the Industry and Startup Levels



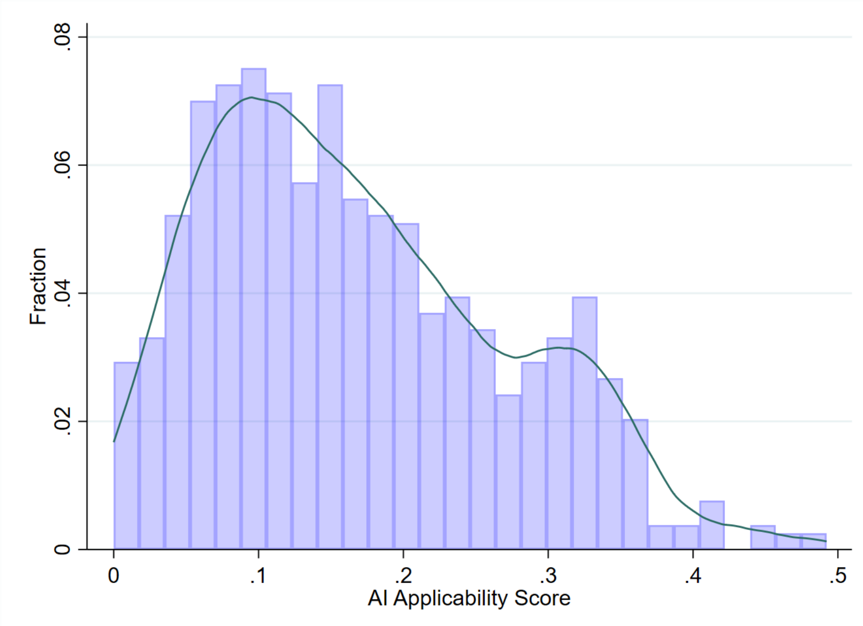
(a) Industry (NAICS 4-digit)



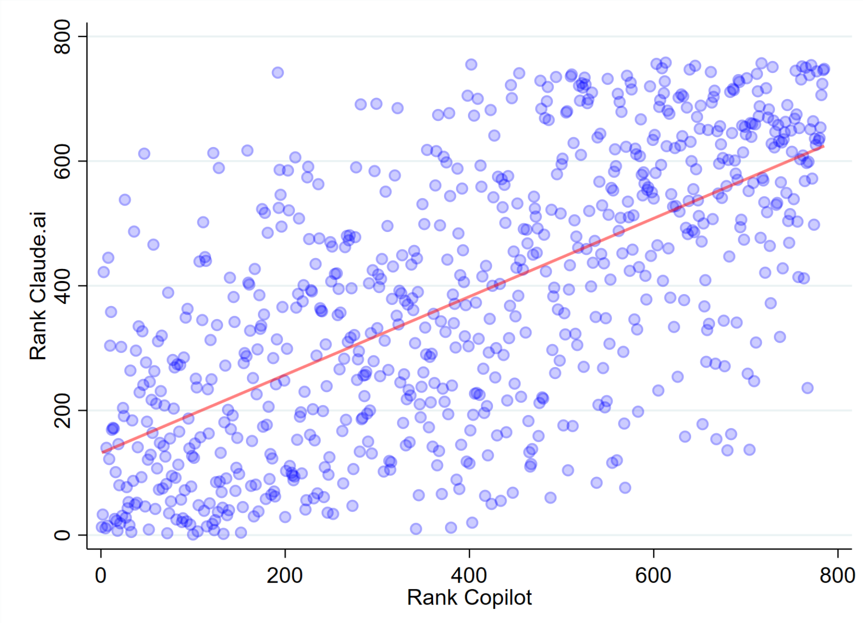
(b) Startup

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 E.2: Exposure Based on Microsoft Bing Copilot



(a) distribution based on Copilot (occupation level)



(b) Claude.ai rank vs. 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.