The Macroeconomics of Generative AI: Firm Entry and Price Dynamics

Jose Carreño*
University of Oregon

March, 2026

Abstract

While the economic literature on generative artificial intelligence has predominantly focused on the disruptive consequences of cognitive automation for the labor market, this paper broadens the analytical lens to product markets. Combining granular occupational AI exposure metrics with administrative records, we document that this technological shift has triggered an immediate, structural surge in firm entry, systematically lowering barriers for high-propensity employer ventures. Concurrently, using a continuous event-study framework, we demonstrate that these firm-level efficiencies pass through to consumers, generating persistent relative price deflation in highly exposed sectors. To rationalize these dual empirical findings, we develop a tractable general equilibrium model of firm heterogeneity. Ultimately, our framework provides a necessary macroeconomic counterweight to prevailing labor-market concerns: by catalyzing structural market dynamism and acting as a potent disinflationary anchor, generative AI generates a substantial and persistent consumer welfare dividend.

JEL Classification: E31, L26, O33
Keywords: Generative AI, Firm Entry, Price Dynamics, General Purpose Technology, Consumer Welfare

*Email: josecarreno.econ@gmail.com.

1 Introduction

Generative Artificial Intelligence (AI) represents one of the most profound technological shifts in modern economic history. Since its widespread commercialization in late 2022, the diffusion of this cognitive technology across the U.S. economy has rapidly accelerated. According to the U.S. Census Bureau's Business Trends and Outlook Survey (BTOS), the share of firms currently utilizing AI to produce goods and services has climbed steadily, while near-term projections remain persistently higher, indicating a strong, continued acceleration in enterprise adoption intent (Figure 1).[1] As this technological diffusion unfolds, both public discourse and the initial wave of economic research have predominantly fixated on the disruptive potential of cognitive automation on the labor market—specifically, the displacement of workers, occupational obsolescence, and the exacerbation of wage inequality (e.g., Autor, 2024).

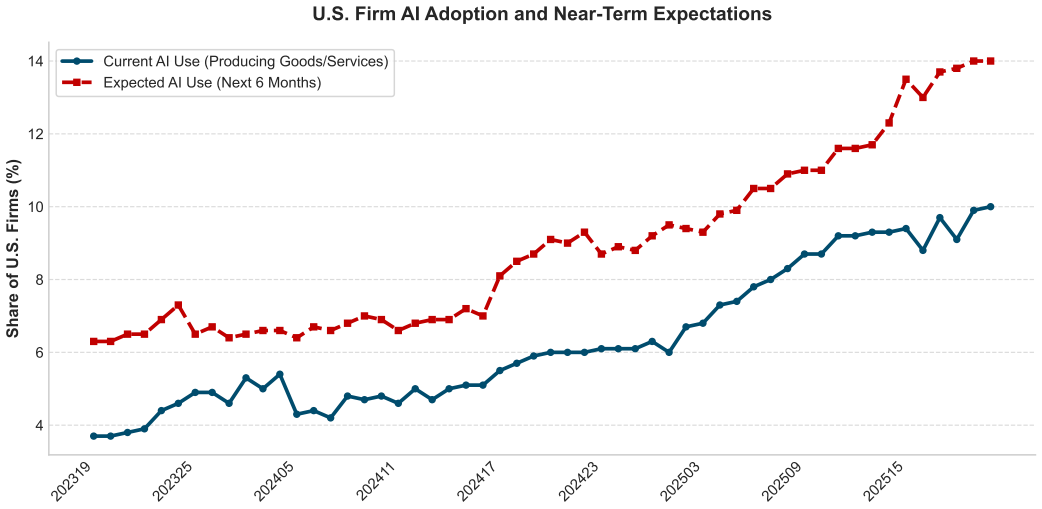


Figure 1: U.S. Firm AI Adoption: Current Use and 6-Month Projections

Notes: The figure plots the bi-weekly share of U.S. firms reporting current use of Artificial Intelligence in producing goods or services, alongside the share of firms anticipating AI usage within the next six months. Data is sourced from the U.S. Census Bureau's Business Trends and Outlook Survey (BTOS).

While the labor-market implications of AI are undeniably profound, this asymmetric focus obscures the broader, general-equilibrium consequences of the technology. From a

[1] In late November 2025, the Census Bureau updated the survey wording from AI use in "producing goods or services" to "in any of its business functions" to better capture auxiliary integration (e.g., accounting, human resources). While this broader definition predictably shifted the reported usage level upward (reaching approximately 18% in early 2026), the current aggregate enterprise adoption rate remains relatively limited. This reinforces the premise that generative AI is still in the early stages of its macroeconomic diffusion.

firm's perspective, generative AI acts as a cognitive General Purpose Technology (GPT) that fundamentally reduces the marginal labor dependencies and fixed administrative overhead required to operate a business (Eloundou et al., 2023). Standard macroeconomic theory posits that such sweeping cost reductions should trigger a dual product-market response: a structural reduction in the barriers to firm entry, and a deflationary downward pressure on consumer prices. In this paper, we construct a unified empirical and theoretical framework to systematically document this transmission. By doing so, we demonstrate that the technological dividends of generative AI extend far beyond pure labor cost savings, ultimately accruing to consumers through heightened market dynamism and enhanced purchasing power.

First, to investigate the transmission of this technological shift into market dynamism, we construct a novel industry-level panel mapping occupational AI exposure to administrative data on U.S. business formations. A fundamental identification challenge in assessing modern entrepreneurial dynamics is the profound, economy-wide upward shift induced by the COVID-19 pandemic, which established a new, elevated baseline of aggregate firm entry. We overcome this confounding factor by constructing a measure of excess firm entry that explicitly re-bases expected counterfactual growth to the post-pandemic "new normal" (2023), isolating the variance that emerged only after the commercial diffusion of generative AI. Our cross-sectional estimates reveal a large and highly significant acceleration in entrepreneurial activity concentrated specifically in highly AI-exposed, knowledge-intensive sectors. Moving from the 25th to the 75th percentile of sectoral AI exposure is associated with a 9.3 percentage point increase in excess firm entry by 2025. Crucially, we document that this technological dividend is not restricted to zero-employee "solopreneurs"; generative AI systematically accelerates the formation of high-propensity ventures with formal payroll structures and explicit hiring intent, confirming its role as a profound democratizing force that structurally lowers barriers to entry.

Second, we evaluate the macroeconomic transmission of this technological shift to product markets and aggregate inflation. If generative AI structurally reduces marginal labor dependencies and fixed overhead costs, standard competitive dynamics dictate that these firm-level efficiencies should inevitably pass through to consumers. To test this disinflationary hypothesis, we exploit a granular panel of sectoral Producer Price Indices (PPI). Because the commercialization of generative AI closely coincided with the extreme volatility of pandemic-era inflation, isolating the technological signal requires rigorous empirical discipline. To this end, we deploy a continuous difference-in-differences (event-study) framework that forces the estimation to navigate the entirety of the COVID-19 supply-chain crisis. The dynamic estimates reveal a striking structural break: while highly exposed and unexposed

sectors followed perfectly parallel price trajectories during the entire pre-pandemic decade, a significant and persistent disinflationary gap emerges post-2022. Crucially, as the transitory inflationary noise of the pandemic fully dissipates by late 2023, this relative price suppression does not revert to the historical mean. Instead, the deflationary gap actively persists and widens through 2025, perfectly mirroring the progressive enterprise integration of AI technologies. This provides compelling evidence that generative AI acts as a potent structural anchor against inflation in highly exposed sectors.

To rationalize these dual empirical findings within a unified framework, we develop a tractable, static general equilibrium model of monopolistic competition with heterogeneous firms, building on the canonical framework of Melitz, 2003. By conceptualizing generative AI as a cognitive General Purpose Technology (GPT) that symmetrically reduces marginal labor dependencies and fixed administrative barriers, the model yields closed-form analytical solutions that formally reconcile our empirical results. Specifically, we demonstrate that the technological reduction in overhead and entry costs mathematically triggers a proportional, sustained influx of new firm varieties. Concurrently, the aggregate price index falls through a dual channel: the direct pass-through of lower marginal costs and the downward competitive pressure exerted by the enhanced product variety. Ultimately, this framework establishes that the combination of structural market dynamism and relative price suppression unambiguously increases aggregate real welfare. By formalizing this substantial and persistent consumer dividend, the model provides a necessary macroeconomic counterweight to the prevailing academic focus on labor displacement, illustrating how the economic gains of cognitive automation are ultimately transmitted to consumers.

It is crucial to contextualize these findings within the broader timeline of technological adoption. As indicated by the current enterprise usage rates (Figure 1), generative AI remains in the nascent stages of its macroeconomic diffusion. Historically, the aggregate integration of General Purpose Technologies—from electrification to the internet—has unfolded over several decades, requiring profound organizational restructuring and the accumulation of new intangible capital (David, 1990; Mokyr, 1990; Brynjolfsson et al., 2021). Consequently, the empirical estimates presented in this paper constitute an early macroeconomic evaluation of an ongoing structural transformation. Because we already identify substantial, statistically robust product-market responses at a relatively limited level of aggregate penetration, our findings should logically be interpreted as a conservative lower bound. As the technology matures, enterprise adoption deepens, and cognitive automation permeates further into complex production processes, the magnitude of this structural dynamism and the resulting consumer dividend are poised to compound significantly over the coming years.

Our paper contributes to three main strands of literature. First, we add a critical macroeconomic perspective to the rapidly expanding literature on the economic impacts of artificial intelligence. To date, this literature has predominantly focused on the labor market implications of cognitive automation, specifically analyzing task displacement, skill-biased technological change, and wage inequality (Acemoglu and Restrepo, 2018; Eloundou et al., 2023; Brynjolfsson et al., 2023; Acemoglu, 2024). While these studies naturally emphasize the disruptive effects of AI on workers and occupational demand, we provide a necessary macroeconomic counterweight by analyzing the transmission of this general purpose technology through product markets. By documenting a structural wave of firm entry and a subsequent relative price deflation in highly exposed sectors, our findings highlight a substantial consumer welfare dividend—driven by enhanced product variety and increased purchasing power. This product-market perspective demonstrates that the technological dividend of generative AI extends far beyond pure labor cost reductions, ultimately accruing to consumers.

Second, our findings contribute to the rich literature on firm dynamics, entrepreneurship, and technological adoption (Hopenhayn, 1992; Decker et al., 2014). Theoretical frameworks have long posited that exogenous reductions in fixed operational and sunk entry costs stimulate business formation and market dynamism. However, empirical evaluations of previous general purpose technologies—such as the ICT revolution—often encountered a modern manifestation of the Solow paradox, where macro-level transformations materialized with significant, multi-decade lags (Brynjolfsson, 1993). In stark contrast, we find that generative AI is acting as an immediate catalyst for economic activity. While recent empirical work has begun to document this rapid technological diffusion accelerating the growth and product innovation of established incumbent firms (Babina et al., 2024), we provide compelling, real-time granular evidence of its immediate impact on *de novo* entrepreneurial activity. By mapping occupational AI exposure directly to the unprecedented surge in high-propensity business applications, we empirically validate the mechanism through which cognitive automation rapidly lowers barriers to entry, fostering a profound wave of structural market dynamism.

Third, our analysis intersects with the macroeconomic literature examining the pass-through of technological innovation to aggregate price dynamics (Oliner et al., 2007; Graetz and Michaels, 2018). While previous studies have documented the disinflationary footprint of earlier automation waves, empirical evidence concerning the specific price effects of modern artificial intelligence remains remarkably scarce. This gap stems from two severe identification challenges: first, the technology is still in the nascent stages of enterprise diffusion;

second, its commercialization directly coincided with the extreme volatility of pandemic-era inflation. We overcome these confounding factors by deploying a continuous event-study design that isolates the underlying technological signal. By doing so, we offer some of the first granular empirical evidence that generative AI acts as a potent disinflationary anchor, even at this early stage of aggregate adoption. Ultimately, our integrated framework suggests that this relative price suppression is the direct theoretical consequence of AI-driven marginal cost reductions and, progressively, enhanced market competition.

The remainder of the paper is organized as follows. Section 2 details the construction of our dataset and the measurement of occupational AI exposure. Section 3 presents the empirical analysis of firm entry. Section 4 evaluates the transmission of this technological shift to product markets and aggregate price dynamics. Section 5 develops our theoretical framework. Finally, Section 6 concludes.

2 Data and Measurement

In this section, we detail the data sources and measurement procedures used to construct our industry-level panels. We combine occupational task evaluations with granular employment and wage data to construct a predetermined, sector-level index of Generative AI exposure. We then pair this index with administrative data on business formations and producer price indices to construct our macroeconomic outcomes. To maintain a clear focus on the core economic mechanisms, we relegate detailed data construction procedures and supplementary descriptive statistics to Appendix A.

2.1 Measuring Generative AI Exposure

To quantify the potential impact of Generative Artificial Intelligence across different sectors of the U.S. economy, we build on the occupational task-level framework developed by Eloundou et al., 2023. Their methodology systematically assesses the extent to which access to a Large Language Model (LLM) reduces the time required for a human worker to perform specific occupational tasks by at least 50 percent. We focus on their β measure, which accounts not only for direct LLM capabilities but also for the broader software and tooling ecosystem built on top of these models. To ensure robustness and mitigate subjective biases, we define our baseline occupational exposure score as the simple average of the human-annotated and machine-generated (GPT-4) evaluations provided in their dataset.

Because our macroeconomic outcomes—firm entry and producer prices—are observed

at the sectoral level, we must map these occupational exposures into an aggregate industry index. Following the standard approach in the literature on task-biased technological change and automation (Autor, 2013; Acemoglu and Restrepo, 2020), we construct the AI exposure of each industry as the wage-bill-weighted average of the exposure of its constituent occupations. Specifically, the AI exposure index for industry j is defined as:

$$AI_Exposure_j = \sum_{o \in j} \left(\frac{w_o E_{o,j}}{\sum_{o \in j} w_o E_{o,j}} \right) Exposure_Score_o \tag{1}$$

where $E_{o,j}$ is the employment of occupation o in industry j, w_o is the mean annual wage of occupation o, and $Exposure_Score_o$ is the occupational AI exposure metric. Weighting by the wage bill, rather than raw employment, appropriately captures the economic cost share of the tasks that are susceptible to AI automation or augmentation within the industry's production function.

We compute this index using granular data from the Bureau of Labor Statistics (BLS) Occupational Employment and Wage Statistics (OEWS) program. Crucially, we utilize the May 2022 OEWS release. By fixing the occupational structure and wage bills to a pre-shock period (prior to the public release of ChatGPT in late 2022), we ensure that our sectoral exposure measure is predetermined and immune to endogenous changes in occupational composition that might have occurred post-AI adoption.

We construct the index at two levels of aggregation to match our outcome variables: the 3-digit NAICS level for the analysis of producer price dynamics, and the broader 2-digit NAICS level for the analysis of business applications.[2]

2.2 Firm Entry: Business Applications

To capture the dynamics of firm entry and entrepreneurial activity, we rely on the Business Formation Statistics (BFS) compiled by the U.S. Census Bureau. The BFS tracks early-stage business formation by aggregating applications for Employer Identification Numbers (EINs) submitted to the Internal Revenue Service. As established by Bayard et al., 2018, these applications serve as a highly reliable, high-frequency, and forward-looking indicator of actual business entry and subsequent job creation in the U.S. economy.

The BFS provides sectoral data aggregated at the 2-digit NAICS level, covering the entire

[2]For the 2-digit NAICS aggregation, we follow the U.S. Census Bureau standard practice of grouping Manufacturing (31-33), Retail Trade (44-45), and Transportation and Warehousing (48-49). See Appendix A for more details.

private sector. A unique advantage of the BFS is its categorization of applications based on their ex-ante likelihood of transitioning into employer businesses (firms with a formal payroll). We exploit this feature by extracting three primary metrics:

1. **Total Business Applications (BA):** The aggregate count of all EIN applications within a sector. We use this as our baseline measure of aggregate entrepreneurial appetite.

2. **High-Propensity Applications (HPA):** A subset of total applications that exhibit characteristics historically associated with a high probability of becoming employer businesses. The Census Bureau flags these applications if they originate from corporate entities, explicitly indicate a date for paying initial wages, or operate in specific industries where non-employer firms are exceptionally rare.

3. **Low-Propensity Applications (LPA):** We define this residual category as the difference between Total BA and HPA. This metric captures the influx of "solopreneurs," independent contractors, and zero-employee startups. Analyzing LPA allows us to isolate a segment of firm entry where barriers consist primarily of bureaucratic setup and cognitive overhead rather than substantial physical capital requirements.

We aggregate the underlying monthly, seasonally adjusted application flows into annual totals to focus on structural shifts rather than high-frequency fluctuations. Our final panel spans the period from 2011 to 2025 across 19 private NAICS sectors, perfectly matching the aggregation level of our Generative AI exposure index. By analyzing this harmonized panel, we can systematically trace how exposure to the AI technological shock affects both the aggregate volume and the compositional nature (HPA vs. LPA) of firm entry.

2.3 Producer Price Indices

To evaluate the pass-through of AI-driven cost reductions into aggregate inflation, our primary outcome variable is the Producer Price Index (PPI) constructed by the Bureau of Labor Statistics (BLS). Unlike the Consumer Price Index (CPI), which incorporates volatile retail margins, imported goods, and consumption taxes, the PPI measures the average change over time in the selling prices received by domestic producers for their output. Because the PPI captures prices at the "factory gate," it is the ideal macroeconomic metric to isolate the direct impact of technology-induced shifts in marginal and fixed operating costs on firm pricing behavior.

To align the price data with our occupational exposure metric, we construct a monthly panel of producer prices at the 3-digit NAICS level. This intermediate level of aggregation offers the optimal balance between securing a sufficiently large cross-sectional sample for statistical identification and minimizing the noise inherent in highly disaggregated product-level price series.

We extract the industry-level PPI series directly from the Federal Reserve Economic Data (FRED) database. While the BLS provides exact 3-digit NAICS price aggregates for the majority of the economy, certain broad sectors are only published at finer levels of disaggregation (e.g., 4-digit or 5-digit NAICS) due to the BLS's hierarchical sampling methodology. To maximize the macroeconomic representativeness of our panel and avoid arbitrary sample selection, we reconstruct the missing 3-digit aggregates by computing the unweighted cross-sectional average of the growth rates of all their surviving sub-sectoral components. This systematic aggregation procedure yields a final, robust panel of 65 distinct 3-digit NAICS sectors spanning the period from January 2011 to December 2025. We relegate detailed data construction procedures to Appendix A.

2.4 Sectoral Characteristics

To account for baseline structural differences across industries that may independently influence firm entry dynamics and pricing behavior, we construct an industry-level measure of labor intensity.

We define the labor share of each sector as the ratio of total employee compensation to total gross domestic product (value added). The data are sourced from the U.S. Bureau of Economic Analysis (BEA) "Components of Value Added by Industry" accounts. To ensure that our covariate captures the fundamental technological structure of the industry—free from the mechanical distortions of the COVID-19 pandemic and strictly predetermined relative to the generative AI adoption—we compute this measure as a multi-year average over the 2016–2019 period:

$$Labor_Share_j = \frac{\sum_{t=2016}^{2019} Compensation_{j,t}}{\sum_{t=2016}^{2019} Value_Added_{j,t}} \tag{2}$$

Consistent with our empirical strategy, we construct this structural covariate at both the 2-digit and 3-digit NAICS levels by aggregating the underlying BEA industry codes to perfectly match the analytical samples used in our subsequent exercises.

3 Firm Entry

The U.S. economy has experienced an unprecedented structural shift in entrepreneurial activity in recent years. Following the onset of the COVID-19 pandemic in early 2020, aggregate business applications increased substantially across virtually all sectors. While this initial, economy-wide surge is widely attributed to pandemic-era macroeconomic factors—such as labor market displacement and shifting consumer preferences—the subsequent evolution of firm entry dynamics in the post-pandemic era reveals substantial cross-sectoral heterogeneity.

Figure 2 illustrates this divergence by plotting the entry dynamics of two representative sectors, selected based on our predetermined Generative AI exposure metric (β). The left panel displays the Professional, Scientific, and Technical Services sector (NAICS 54), which ranks among the most highly exposed sectors in our sample ($\beta = 0.52$, implying that 52% of the sectoral wage bill is allocated to AI-susceptible occupational tasks). Following the initial pandemic shock, business formations in this knowledge-intensive sector remained persistently elevated, with the upward trend steepening progressively in the quarters following the public release of generative AI models in late 2022. Conversely, the right panel plots Accommodation and Food Services (NAICS 72), a sector characterized by a low AI exposure score ($\beta = 0.20$). While NAICS 72 also experienced a significant upward structural break during the pandemic, its entry rate trend subsequently stabilized, showing no further acceleration over the 2023–2025 period.

While these two sectors serve as illustrative archetypes, this diverging pattern is not idiosyncratic. Rather, the underlying systematic divergence provides a valuable empirical laboratory to study the macroeconomic implications of this new technological adoption. It suggests that while the pandemic induced an economy-wide level shift in firm entry, the subsequent diffusion of generative AI may be systematically lowering the barriers to entry in specifically exposed industries. In this section, we formalize this hypothesis. By exploiting our full cross-sectoral panel, we isolate the variance in entry dynamics post-2022 to examine the formal relationship between generative AI exposure and both the aggregate volume and compositional nature of business formations.

3.1 Empirical Strategy

The primary objective of our empirical analysis is to systematically evaluate whether the widespread adoption of generative AI has lowered the barriers to firm entry. Theoretically, generative AI functions as a cognitive, labor-saving technology. By automating or

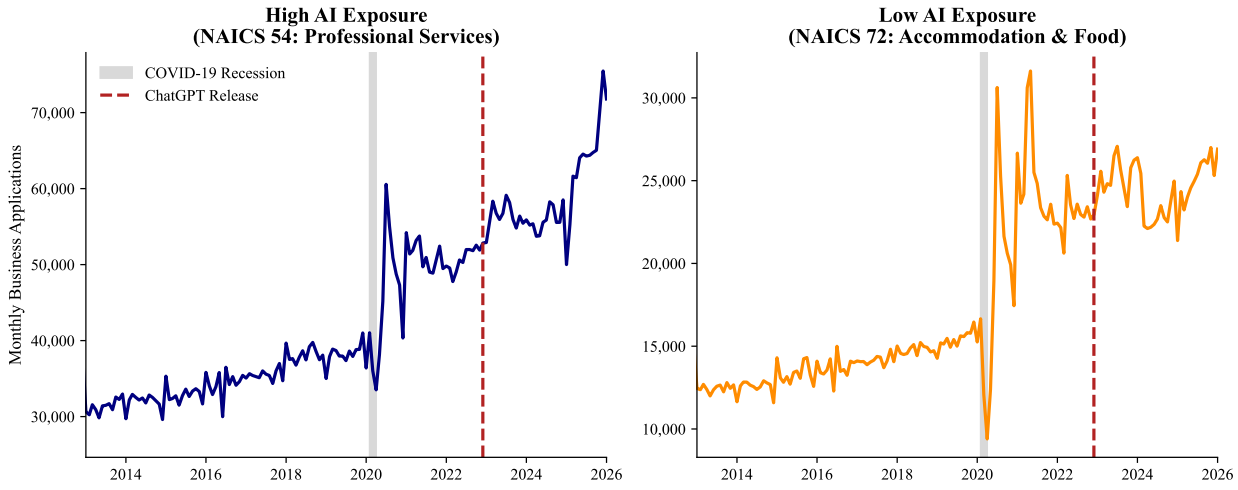


Figure 2: Business Applications by Generative AI Exposure (2013–2026)

Notes: The figure contrasts the monthly, seasonally adjusted Total Business Applications for a high AI-exposure sector (NAICS 54: Professional, Scientific, and Technical Services) and a low AI-exposure sector (NAICS 72: Accommodation and Food Services). The shaded gray area denotes the COVID-19 recession as defined by the NBER (February 2020 to April 2020). The vertical dashed red line marks the public release of ChatGPT in November 2022. Data source: U.S. Census Bureau, Business Formation Statistics.

augmenting tasks that previously required human capital, it significantly reduces the fixed bureaucratic overhead and marginal payroll costs associated with establishing and scaling a new business. If these cognitive and labor requirements constitute a binding constraint on entrepreneurship, industries with higher exposure to this technology should experience a disproportionate acceleration in firm entry.

Identifying this relationship in aggregate data, however, presents two fundamental econometric challenges. First, U.S. industries exhibit inherent, secular differences in their entry dynamics; knowledge-intensive sectors naturally display different baseline growth trends than traditional manufacturing or retail sectors, irrespective of any technological shock. Second, the COVID-19 pandemic induced a profound, economy-wide structural break, establishing a new, elevated baseline of entrepreneurial activity that confounds simple pre- versus post-AI comparisons.

To overcome these identification challenges, we construct a measure of *excess firm entry* that explicitly accounts for both sector-specific historical growth trends and the post-pandemic level shift. For our baseline specification, we first estimate a log-linear growth trend for each sector j using the pre-pandemic period (2015–2019):

$$\ln(BA_{j,t}) = \alpha_j + \delta_j t + \varepsilon_{j,t} \tag{3}$$

We then use the estimated trend parameter ($\hat{\delta}_j$) to extract the expected counterfactual growth factor for the post-AI period. Crucially, rather than projecting this trend forward from 2019—which would erroneously conflate the AI effect with the pandemic-induced structural break—we re-base our projection to the actual observed volume of business applications in 2023 ($BA_{j,2023}$).

We select 2023 as our baseline year because it is sufficiently removed from the acute disruptions of the pandemic, thereby capturing the "new normal" after the upward shift. While the public release of ChatGPT occurred in late 2022, enterprise-level adoption and business formation inherently involve time lags. Therefore, setting 2023 as the base is a conservative methodological choice: any early, AI-induced firm entry that occurred during 2023 is absorbed into the baseline, meaning our estimates will likely represent a lower bound of the true technological effect.

Using this framework, we compute the counterfactual expected entry in 2025 as $\tilde{BA}_{j,2025} = BA_{j,2023} \times \exp(2\hat{\delta}_j)$. We then define our dependent variable, the normalized *Excess Entry Gap*, as the difference between the observed and counterfactual applications in 2025, scaled by the 2023 baseline:

$$Excess_BA_j = \frac{BA_{j,2025} - \tilde{BA}_{j,2025}}{BA_{j,2023}} \tag{4}$$

Finally, to isolate the effect of AI exposure from broader structural characteristics, we incorporate the industry's historical labor share as a fundamental control. The theoretical rationale is straightforward: if AI primarily reduces labor requirements, its capacity to induce firm entry will be constrained in highly capital-intensive sectors where substantial physical investments remain the primary barrier to entry. Conversely, in highly labor-intensive sectors, the AI-driven reduction in cognitive and bureaucratic costs may remove the only binding barrier to entry, leading to a significantly stronger entrepreneurial response.

We formalize our empirical test via the following cross-sectional regression across our 19 private NAICS sectors:

$$Excess_BA_j = \beta_0 + \beta_1 AI_Exposure_j + \gamma Labor_Share_j + u_j \tag{5}$$

where $AI_Exposure_j$ is our predetermined occupational exposure index. The model is estimated using heteroskedasticity-robust standard errors. To ensure that our findings are not driven by arbitrary methodological choices, we complement this baseline specification with an extensive battery of robustness checks, detailed later in the text. These include alternative pre-trend horizons (e.g., 2011–2019), quadratic trend specifications, direct growth-rate

regressions controlling for pre-trends, and leave-one-out cross-validation.

To interpret these estimates as the causal effect of generative artificial intelligence, it is imperative to consider potential threats to our identification strategy. The primary econometric challenge stems not from aggregate macroeconomic shocks, but from unobserved confounders that might have differentially affected firm entry dynamics while simultaneously correlating with our measure of AI exposure. It is conceptually useful to separate these potential biases into demand-side and supply-side forces.

On the demand side, a natural concern is that the COVID-19 pandemic induced a permanent structural shift in consumer and business preferences, generating a sustained surge in demand for knowledge-intensive services. If this were the case, the excess firm entry we observe would not reflect a technology-driven reduction in entry barriers, but rather a classic pull-effect driven by expanding profit margins in those sectors. However, the empirical viability of this demand-side counterfactual is naturally bounded by the behavior of prices. As we demonstrate in Section 4, sectors with higher AI exposure do not experience inflationary pressures; conversely, they exhibit a marked downward trend in their relative prices. This combined dynamic—heightened firm entry coupled with sectoral price deflation—points unequivocally to an expansionary supply shock, largely ruling out demand as the primary driver of the phenomenon.

On the supply side, the most prominent disruptive force (apart from AI itself) has been the widespread transition to remote work. Undeniably, the ability to operate remotely has facilitated business formation in knowledge-intensive industries by reducing the need for physical infrastructure. Nevertheless, the shift to remote work was an abrupt shock that materialized and was largely internalized between 2020 and 2021. By re-basing our counterfactual projections using the observed entry volume in 2023, the level effects of this new labor paradigm are absorbed into our baseline. For remote work, or any isomorphic unobserved factor, to bias our results, its capacity to reduce entry barriers would have had to experience a second, sudden, and silent acceleration exactly in late 2022—a scenario that lacks empirical support. Furthermore, the impact of other capital-saving technologies is likely mitigated by the inclusion of our structural labor intensity control.

Ultimately, our empirical strategy rests on the assumption that pre-pandemic trend growth rates (rather than their intercepts) provide a reasonable approximation of the counterfactual trajectory of the economy post-2023.

Overall, while no macroeconomic empirical exercise is entirely immune to endogeneity concerns, the requirement that an unobserved confounder must simultaneously (i) strongly correlate with the susceptibility to cognitive task automation, (ii) suddenly accelerate after

2022, and (iii) generate both a firm entry boom and sectoral deflation, drastically restricts the set of plausible alternative explanations.

3.2 Baseline Results

We begin by providing visual evidence of the unconditional relationship between generative AI exposure and subsequent firm entry dynamics. Figure 3 illustrates the cross-sectional correlation between our occupational exposure index (β) and the normalized excess business applications in 2025, computed using the short pre-pandemic trend (2015–2019). We observe a tight and strongly positive linear relationship. Notably, the distribution lacks asymmetric observations that would contradict our theoretical premise: the quadrants corresponding to low exposure with high excess entry, or high exposure with low excess entry, are empty. While the mechanical levels of the excess gap tend to be negative on average—a mathematical artifact of projecting steep pre-pandemic growth rates onto the structurally elevated 2023 base—the relative cross-sectional variation clearly indicates that highly exposed sectors systematically outperform their low-exposure counterparts relative to their historical trajectories.

Table 1 formalizes this visual evidence through our baseline cross-sectional regressions. Column (1) reports the estimates using our preferred 2015–2019 trend projection, which most accurately captures the growth dynamics immediately preceding the pandemic. Column (2) provides complementary evidence using a longer historical trend (2011–2019). In both specifications, we include the sector's historical labor share to isolate the technological effect from underlying factor intensities.

The point estimates confirm a highly significant and economically substantial relationship. In our preferred specification (Column 1), the coefficient on generative AI exposure is 0.731 (significant at the 1% level). To contextualize this economic magnitude, we evaluate the effect in terms of the interquartile range of sectoral exposure (documented in Table A.1). Moving from the 25th percentile ($\beta = 0.296$) to the 75th percentile ($\beta = 0.423$) of the technological distribution is associated with a 9.3 percentage point increase in excess firm entry relative to the 2023 baseline. This effect accounts for a substantial fraction of the cross-sectoral divergence observed in the post-pandemic recovery. Furthermore, the inclusion of the labor intensity control empirically validates our theoretical intuition; its positive and statistically significant coefficient (0.326) confirms that sectors inherently dependent on labor rather than large physical investments experience structurally lower barriers to entry.

Indeed, this parsimonious two-variable model explains 56% of the total cross-sectional

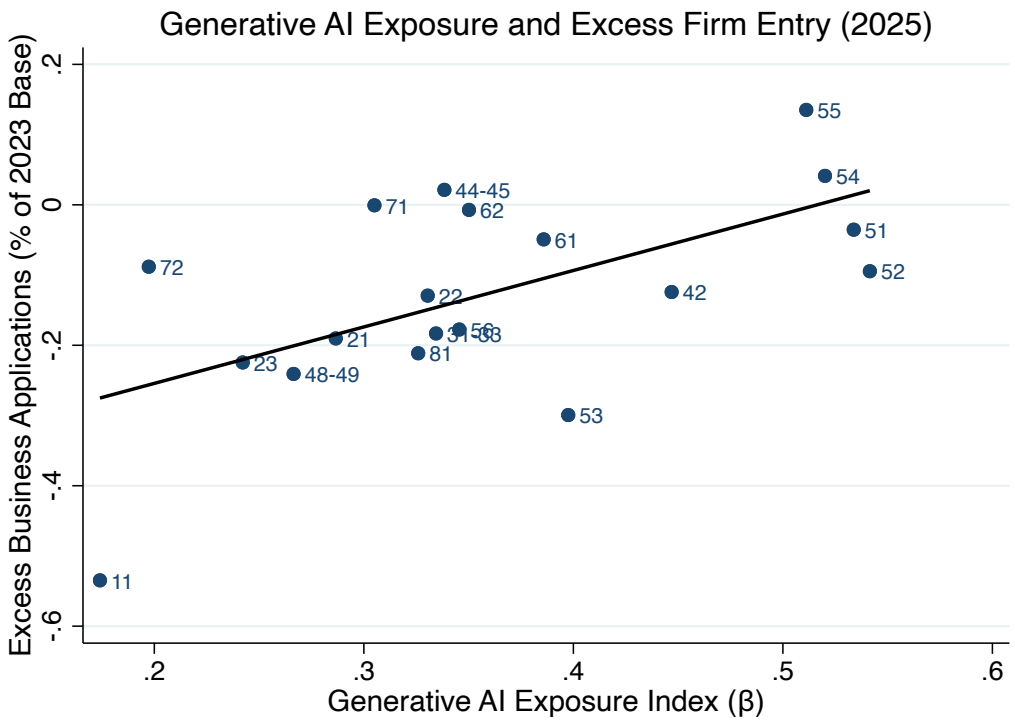


Figure 3: Generative AI Exposure and Excess Firm Entry

Notes: The figure presents the cross-sectional scatter plot and linear fit between the Generative AI occupational exposure index (β) and the *Excess Entry Gap* for 19 private NAICS sectors. The excess gap is defined as the difference between observed 2025 business applications and the counterfactual projection (based on the 2015–2019 trend), normalized by the 2023 base volume.

variance in post-2023 excess entry dynamics ($R^2 = 0.560$). Given the limited statistical power inherent in a sample of only 19 sectoral observations, this level of explanatory power strongly underscores the macroeconomic relevance of this technological shift. To ensure that this high level of significance is not spuriously driven by influential outliers or by artifacts in the normative construction of the dependent variable, Section 3.4 subjects these baseline estimates to an extensive battery of empirical robustness checks. These exercises include leave-one-out cross-validation, the systematic exclusion of broader industry groups, quadratic trend specifications, and direct growth-rate estimations.

3.3 Compositional Heterogeneity: High vs. Low Propensity

Building on the baseline relationship between generative AI exposure and aggregate firm entry, a natural subsequent question pertains to the composition of these new ventures. Specifically, we investigate whether this technological shift is predominantly fueling a surge

Table 1: Baseline Results: Generative AI Exposure and Firm Entry

	(1) Short Trend (2015-2019)	(2) Long Trend (2011-2019)
Generative AI Exposure (β)	0.731***	0.809***
	(0.243)	(0.254)
Labor Share (2016-2019)	0.326***	0.299**
	(0.089)	(0.106)
Observations	19	19
R^2	0.559	0.531

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the normalized *Excess Entry Gap* in 2025, computed using the 2015–2019 pre-trend in Column (1) and the 2011–2019 pre-trend in Column (2). The Generative AI Exposure (β) and Labor Share are measured as pre-determined cross-sectional characteristics at the NAICS 2-digit level. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

of independent contractors and zero-employee "solopreneurs," or if it is also systematically reducing barriers for structurally complex firms that intend to hire and scale.

To explore this compositional heterogeneity, we leverage the Census Bureau's ex-ante categorization of business formations. As outlined in Section 2, High-Propensity Applications (HPA) identify ventures with characteristics historically linked to a high probability of transitioning into employer businesses (e.g., corporate entities or firms providing a specific date for initial payroll). Conversely, Low-Propensity Applications (LPA) capture the residual influx of independent contractors and initial zero-employee startups. We replicate our baseline empirical strategy separately for both the HPA and LPA components.

Table 2 reports the results. The estimates reveal a pervasive, across-the-board effect: generative AI exposure significantly accelerates firm entry across both categories.

The strong response among High-Propensity Applications (Columns 1 and 2) is particularly noteworthy from a macroeconomic perspective. It demonstrates that the AI-induced entrepreneurial expansion is not restricted solely to non-employer micro-businesses. Instead, the results indicate that the technology is actively lowering the barriers for ventures with explicit hiring intent and formal payroll structures. Consistent with our hypothesis, by reducing the overhead costs associated with administrative, legal, and cognitive tasks, AI facilitates the establishment of more complex firms that are crucial for aggregate job creation.

Concurrently, the estimates for Low-Propensity Applications (Columns 3 and 4) are

Table 2: Heterogeneity Analysis: High vs. Low Propensity Firm Entry

	High-Propensity (HPA)		Low-Propensity (LPA)	
	(1) Short Trend	(2) Long Trend	(3) Short Trend	(4) Long Trend
Generative AI Exposure (β)	0.639***	0.670**	0.875***	0.954***
	(0.197)	(0.249)	(0.292)	(0.289)
Labor Share (2016-2019)	0.288**	0.266	0.317***	0.298**
	(0.103)	(0.153)	(0.102)	(0.105)
Observations	19	19	18	18
R^2	0.554	0.403	0.502	0.513

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the normalized *Excess Entry Gap* for High-Propensity Applications (Columns 1-2) and Low-Propensity Applications (Columns 3-4). The sample size for the LPA specifications is 18 due to a Census Bureau methodological feature that automatically categorizes all applications in the Accommodation and Food Services sector (NAICS 72) as High-Propensity. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

highly significant and nominally larger in magnitude.[3] This aligns seamlessly with the theoretical premise that generative AI serves as a powerful democratizing force in entrepreneurship. Consistent with a reduction in initial cognitive and bureaucratic friction, the technology enables individuals to launch zero-employee ventures that previously might have carried prohibitive fixed costs. While many of these LPA entities may remain small, the literature on firm dynamics emphasizes that a fraction of these initial solopreneurs will eventually scale (Haltiwanger et al., 2013; Decker et al., 2014), making this expansive entry margin a critical component of future economic dynamism.

3.4 Robustness Checks

To ensure that our findings are not driven by sample idiosyncrasies or specific functional form assumptions, we subject our baseline estimates to an extensive series of robustness exercises. Full estimation results for these tests are provided in Appendix B.

First, we address potential concerns regarding sample composition. Given the constrained statistical power inherent in a cross-sectional sample of $N = 19$ aggregate sectors,

[3]While the point estimates for LPA are nominally larger than those for HPA, standard Wald tests indicate that the coefficients are not statistically distinguishable from each other at conventional significance levels. This reinforces the interpretation of a generalized reduction in entry barriers across the entrepreneurial spectrum.

parameter estimates could theoretically be sensitive to influential outliers. To systematically rule out the possibility that the aggregate effect is spuriously driven by any single industry, we perform a leave-one-out (jackknife) sensitivity analysis.

Panel A of Appendix Table B.1 demonstrates that the coefficient on AI exposure remains remarkably stable and statistically significant (at the 1% or 5% level) across all 19 iterative exclusions. To impose a strictly more demanding test, Panel B concurrently excludes pairs of knowledge-intensive sectors (e.g., jointly dropping Information and Professional Services). Even after artificially truncating these concentrated clusters, the technological exposure coefficient remains strongly stable, confirming that the systematic reduction in entry barriers is a generalized, economy-wide phenomenon rather than an artifact of a specific professional cluster.

Second, we verify that our results are not artifacts of the methodological construction of the dependent variable. In our baseline specification, we define the outcome as the normalized gap relative to a log-linear trend. Appendix Table B.2 demonstrates that our main findings remain remarkably similar when we replace the log-linear projection with a more flexible quadratic historical trend (Column 1), which naturally accommodates non-linear pre-pandemic acceleration paths.

More stringently, we dispense with the gap-based contrafactual entirely and estimate direct growth-rate regressions. In Columns 2 and 3 of Appendix Table B.2, the dependent variable is simply the observed post-AI sectoral growth rate (the direct log difference from 2023 to 2025). To ensure robustness against the choice of specific base years, Columns 4 and 5 evaluate the growth rate between the 2022–2023 average and the 2024–2025 average. In all specifications, we directly control for the sector's pre-pandemic compound annual growth rate (CAGR) on the right-hand side of the regression. Regardless of whether we use the short or long pre-trend to compute the historical CAGR control, the AI exposure coefficient remains large, positive, and highly significant. This straightforward empirical exercise confirms that even when evaluated using raw growth differentials, sectors highly exposed to cognitive automation have accelerated significantly faster than their historical baselines.

Finally, while this cross-sectional framework cleanly isolates the persistent variance in aggregate entry rates, the coarse aggregation of our 19-sector panel inherently restricts the feasibility of a formal event-study design. We overcome this limitation in Section 4, where we are able to exploit a significantly more granular industry classification for sectoral prices.

4 Price Dynamics

The persistent acceleration of firm entry documented in Section 3 firmly establishes that highly AI-exposed sectors are experiencing profound structural dynamism. However, the ultimate macroeconomic implications of this technological shift fundamentally depend on its transmission to product markets and consumer prices.

Our core hypothesis posits that as generative AI automates and augments cognitive tasks, it structurally reduces labor dependencies and marginal operating costs. Under standard competitive market conditions, where prices reflect a markup over marginal costs, these firm-level efficiencies should inevitably pass through to consumers. Consequently, sectors with high exposure to cognitive automation should experience significant downward pressure on their relative prices. This dynamic is macroeconomically crucial: if this new general-purpose technology acts as a structural disinflationary anchor, the combination of increased market entry and relative price deflation yields a substantial dividend for consumer welfare.

To formalize this hypothesis, we exploit a granular panel of sectoral Producer Price Indices (PPI) from the Bureau of Labor Statistics. For consistency across our empirical designs, we aggregate the underlying monthly price indices into annual averages to focus strictly on structural shifts rather than high-frequency fluctuations. We then evaluate the disinflationary impact of generative AI through two independent but highly complementary empirical strategies.

First, we estimate cross-sectional inflation gaps relative to historical pre-pandemic trajectories. By projecting the counterfactual price levels from 2019 directly to 2025, this cross-sectional approach effectively steps over the extreme, transitory noise of the pandemic-era supply chain disruptions, allowing us to cleanly isolate the cumulative price divergence.

Second, we subject our hypothesis to a strictly more demanding empirical test by implementing a dynamic event-study (continuous difference-in-differences) framework. Utilizing the full annual panel, this approach does not bypass the pandemic period; rather, it forces the model to navigate the entirety of the COVID-19 inflationary shock. This demanding design allows us to rigorously test for the presence of parallel pre-trends and precisely map the dynamic pass-through of the AI technological shock following its widespread commercialization. Together, these two independent frameworks provide a robust empirical foundation to evaluate the disinflationary footprint of artificial intelligence.

4.1 Cross-Sectoral Inflation Gaps

To formally test the disinflationary hypothesis, our first empirical exercise evaluates the cumulative price divergence across sectors relative to their pre-pandemic trajectories. Consistent with our approach to firm entry, we construct a measure of excess inflation that accounts for secular, sector-specific price dynamics.

We begin by estimating the baseline inflation trend for each sector j over the pre-pandemic period (2015–2019). Given the high degree of persistence characteristic of price indices, we model the log-level of the Producer Price Index (PPI) as a function of time:

$$\ln(PPI_{j,t}) = \alpha_j + \delta_j t + \varepsilon_{j,t} \tag{6}$$

Using the estimated parameters, we project the counterfactual price trajectory for the post-AI period. Specifically, the expected price index in 2025—representing the counterfactual price level that would have prevailed had the pandemic and subsequent technological shocks not occurred—is given by $\tilde{PPI}_{j,2025} = \exp(\hat{\alpha}_j + \hat{\delta}_j t_{2025})$. We then define our primary dependent variable, the *Inflation Gap*, as the percentage deviation of the actual observed prices in 2025 from this baseline counterfactual:

$$Inflation_Gap_j = \frac{PPI_{j,2025} - \tilde{PPI}_{j,2025}}{\tilde{PPI}_{j,2025}} \tag{7}$$

This metric captures the cumulative excess inflation absorbed by each sector. If generative AI acts as a deflationary force by reducing marginal costs, sectors with higher occupational exposure should exhibit systematically smaller—or even negative—inflation gaps relative to unexposed sectors.

Evaluating this cumulative gap at the end of our sample period in 2025 serves a dual methodological purpose. First, the commercialization and enterprise adoption of generative AI models is a progressive, cumulative process. Measuring the gap in 2025 captures the maximum extent of technological diffusion available in current data. Second, and crucially for identification, evaluating price divergence several years after the acute phase of the pandemic minimizes the confounding influence of this transitory macroeconomic shock. By 2024 and 2025, the severe supply chain bottlenecks and demand-supply imbalances that characterized the immediate post-pandemic period had largely normalized.[4] Evaluating prices at this

[4]This macroscopic normalization is empirically well-documented. The Federal Reserve Bank of New York's Global Supply Chain Pressure Index (GSCPI)—which peaked at an unprecedented 4.3 standard deviations above its historical mean in late 2021—fully reverted to its historical average by February 2023

remote point provides confidence that the observed cross-sectoral variation reflects persistent structural shifts rather than the lingering asymmetric noise of the pandemic era.

Before introducing our formal econometric specification, Figure 4 provides visual evidence of this unconditional relationship. The scatter plot displays a clear, negative cross-sectional correlation between our occupational AI exposure index (β) and the excess inflation gap in 2025. While the sample size here is substantially more granular than in our firm entry analysis, the underlying pattern remains striking: sectors highly exposed to cognitive automation systematically exhibit lower excess inflation relative to their historical baselines.

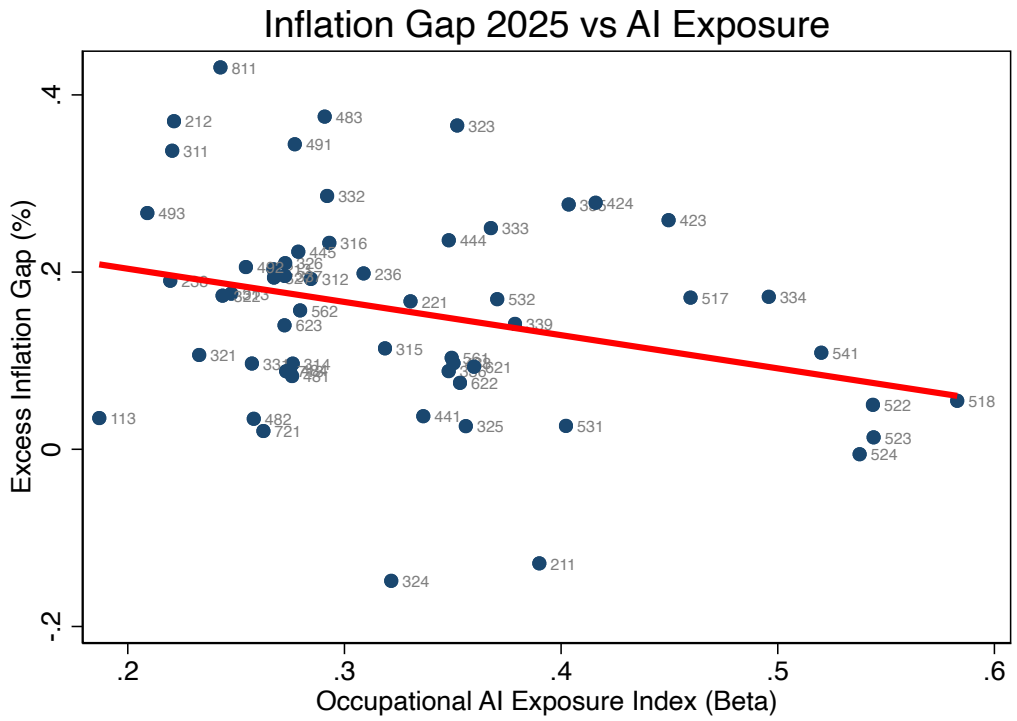


Figure 4: Generative AI Exposure and Sectoral Inflation Gaps

Notes: The figure presents the cross-sectional scatter plot and linear fit between the Generative AI occupational exposure index (β) and the *Excess Inflation Gap* for our granular sample of NAICS industries. The gap is defined as the percentage deviation of the observed 2025 Producer Price Index from its counterfactual projection based on the short pre-pandemic trend (2015–2019).

To isolate the technological effect from fundamental differences in production factor intensities, we retain the sector's historical labor share as a structural control. The mi-

(Federal Reserve Bank of New York, 2023). Corroborating this timeline, the Federal Open Market Committee formally assessed in late 2023 that the healing of global supply chains and labor supply imbalances was largely complete (Federal Open Market Committee, 2023). Consequently, the 2024–2025 period provides a clean empirical window where transitory supply bottlenecks no longer dictate relative price dynamics.

croeconomic rationale is straightforward: in highly capital-intensive sectors, AI-driven labor savings represent a relatively small fraction of total marginal costs, naturally limiting the potential for significant downward price pass-through. This control ensures that our estimates are not confounded by the general divergence in price dynamics between capital- and labor-intensive sectors during the post-pandemic recovery. We formalize this cross-sectional relationship through the following specification:

$$Inflation_Gap_j = \gamma_0 + \gamma_1 AI_Exposure_j + \theta Labor_Share_j + v_j \tag{8}$$

where γ_1 is our coefficient of interest, expected to be negative. To ensure our estimates are robust to the functional form and the horizon of the baseline projection, we estimate Equation (3) using the baseline 2015–2019 trend, a longer 2011–2019 linear trend, and a more flexible quadratic pre-trend that accommodates potential non-linearities in pre-pandemic price acceleration.

Table 3 formalizes this visual evidence, reporting the cross-sectional estimates of Equation (3). Across all specifications, the coefficient on AI exposure is negative and statistically significant, confirming a robust disinflationary pass-through in highly exposed sectors.

Column (1) presents our preferred specification, which defines the counterfactual projection based on the most recent pre-pandemic trend (2015–2019). The point estimate of -0.326 is significant at the 5% level. To contextualize the economic magnitude of this deflationary pressure, we evaluate the effect across the interquartile range of our 3-digit NAICS sample (see Table A.1). Moving from a sector at the 25th percentile of technological exposure ($\beta = 0.273$) to one at the 75th percentile ($\beta = 0.402$) is associated with a 4.2 percentage point reduction in the cumulative excess inflation gap.

This relative price suppression is remarkably robust to alternative functional forms of the underlying price trajectory. Column (2) extends the linear pre-trend to the 2011–2019 period, while Column (3) employs a flexible quadratic trend to allow for non-linear historical acceleration. In both alternative specifications, the point estimates for the AI exposure parameter become larger in magnitude (-0.445 and -0.520, respectively) and remain statistically significant, suggesting that our baseline specification provides a conservative lower bound of the disinflationary effect.

While these cross-sectional results provide compelling evidence of cumulative relative price suppression, estimating gaps inherently relies on the validity of the chosen counterfactual projections. Furthermore, measuring outcomes at a single point in time (2025) obscures the dynamic realization of these price adjustments. To address these limitations and conclu-

Table 3: Cross-Sectoral Estimates: Generative AI and Inflation Gaps

	(1) Short Trend	(2) Long Trend	(3) Quadratic Trend
Generative AI Exposure (β)	-0.326**	-0.445**	-0.520**
	(0.148)	(0.219)	(0.233)
Labor Share (2016-2019)	0.126	-0.513*	0.252
	(0.092)	(0.286)	(0.161)
Observations	57	57	57
R^2	0.128	0.156	0.149

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the *Excess Inflation Gap*, defined as the percentage deviation of the observed 2025 Producer Price Index from its historical counterfactual projection. Column (1) projects from the 2015–2019 linear trend; Column (2) from the 2011–2019 linear trend; and Column (3) from the 2011–2019 quadratic trend. The estimation sample comprises 57 NAICS 3-digit industries due to the limited availability of historical labor share data for a small subset of 6 sectors. Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sively identify the temporal pass-through of the technological effect, we turn to our dynamic event-study design.

4.2 Dynamic Event Study

To conclusively identify the dynamic pass-through of the technological effect and formally test the internal validity of our empirical design, we implement a generalized event-study (continuous difference-in-differences) framework.

We estimate the following two-way fixed effects specification:

$$\ln(PPI_{j,t}) = \alpha_j + \tau_t + \sum_{k=2011, k\neq 2019}^{2025} \beta_k \left(AI_Exposure_j \times \mathbb{I}\{year = k\}\right) + \varepsilon_{j,t} \tag{9}$$

where $\ln(PPI_{j,t})$ is the log price index for sector j in year t. The model includes sector fixed effects (α_j) to absorb time-invariant unobserved heterogeneity across industries—effectively subsuming static structural controls such as historical labor share—and year fixed effects (τ_t) to control for aggregate macroeconomic shocks, including the baseline inflationary waves of the pandemic.

The coefficients of interest, β_k, capture the dynamic, year-specific price differential associated with a one-unit higher exposure to generative AI, relative to the pre-pandemic base year

of 2019. Following standard practice in continuous treatment designs (Autor et al., 2006), the validity of our identifying assumption rests on the absence of pre-trends; the estimated β_k coefficients should be statistically indistinguishable from zero prior to 2020. Standard errors are clustered at the NAICS industry level to allow for arbitrary within-sector autocorrelation over time.

This dynamic panel approach represents a highly demanding empirical test for two distinct reasons. First, unlike the cross-sectional gap strategy which circumvents the 2020–2023 period, the event study forces the model to continuously navigate the extreme volatility of the pandemic and the subsequent supply-chain crisis. Consequently, demonstrating a stable, parallel pre-trend prior to 2020, followed by a sustained structural break post-2022, serves as a stringent test to rule out the concern that our results are driven by lingering, asymmetric COVID-19 disruptions.

Second, evaluating a continuous treatment effect within a two-way fixed effects framework inherently absorbs a vast amount of empirical variation. Given the relatively compact cross-sectional dimension of our granular sample ($N = 65$), statistical power is mechanically constrained. Moreover, as established in the econometric literature, the introduction of unit fixed effects in the presence of any classical measurement error in the exposure index exacerbates attenuation bias, systematically driving point estimates toward zero (Griliches and Hausman, 1986). Consequently, successfully identifying a statistically significant dynamic effect in this framework provides strong reassurance regarding the robustness of the signal, with the resulting estimates arguably representing a conservative lower bound of the true disinflationary impact.

Figure 5 plots the estimated dynamic coefficients (β_k) alongside their 90% and 95% confidence intervals. The results provide striking visual and statistical confirmation of our empirical design. Most crucially, the point estimates prior to 2020 are tightly clustered around zero and lack any secular trajectory. This absence of pre-trends rigorously validates our identifying assumption: highly AI-exposed sectors and unexposed sectors were on parallel price trajectories during the entire pre-pandemic decade, ruling out the presence of confounding, long-run structural drifts.

Following the onset of the pandemic, the temporal dynamics reveal a fascinating interplay between two distinct macroeconomic forces. We observe an initial downward divergence in relative prices somehow starting in 2021 and becoming statistically significant by 2022. While this early structural break loosely coincides with the nascent introduction of generative AI tools, a conservative empirical interpretation must acknowledge the concurrent, asymmetric effects of the pandemic. Highly AI-exposed industries—predominantly knowledge-intensive

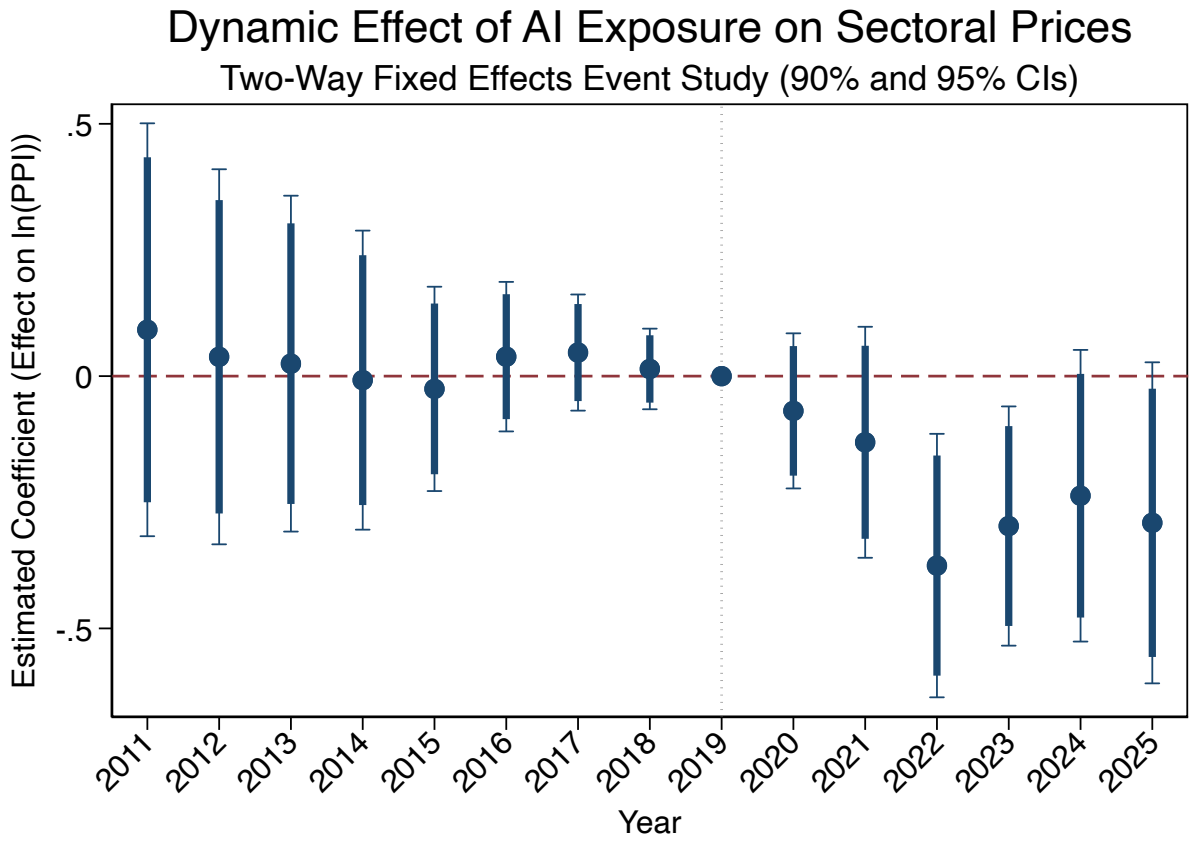


Figure 5: Dynamic Effect of Generative AI Exposure on Sectoral Prices

Notes: The figure plots the β_k coefficients from the continuous difference-in-differences specification (Equation 9), representing the dynamic effect of a one-unit increase in the Generative AI exposure index (β) on the log Producer Price Index. The year 2019 serves as the omitted base period. The inner (thick) and outer (thin) vertical lines represent the 90% and 95% confidence intervals, respectively. Standard errors are clustered at the NAICS industry level. Full estimation results are provided in Appendix B.1.

service sectors—were naturally more insulated from the severe physical supply-chain bottlenecks and commodity shocks that disproportionately inflated prices in goods-producing sectors during the acute phase of the crisis.

However, the true strength of the technological hypothesis emerges in the subsequent years. If this initial divergence were purely an artifact of pandemic-era supply bottlenecks, standard macroeconomic mean-reversion would dictate a rapid closing of the gap by 2023 and 2024, precisely after global supply chains fully normalized. Nevertheless, the point estimates exhibit a slight stabilization during this intermediate period, coinciding with the fading of the pandemic noise. Yet, rather than reverting to the pre-2020 baseline, the disinflationary gap actively persists and begins to widen again as we approach 2025.

This sustained and deepening structural break perfectly mirrors the progressive commercialization, enterprise integration, and compounding efficiency gains of generative AI models. As the transitory pandemic shock dissipates, the technological effect unambiguously takes over, cementing a permanent downward shift in the relative price trajectory of highly exposed sectors.

To further explore the frontier of this dynamic, Appendix B.1 extends the event-study estimation to incorporate the latest available data point (January 2026). While interpreting a single-month observation warrants statistical caution due to the absence of full-year smoothing, the point estimate drops markedly further. This provides compelling, suggestive evidence that the deflationary pressure exerted by artificial intelligence is not only persistent, but actively compounding as technological diffusion deepens across the aggregate economy.

5 Theoretical Framework

To rationalize our empirical findings—namely, that exposure to generative AI induces a sustained wave of firm entry (Section 3) while simultaneously exerting downward pressure on sectoral prices (Section 4)—we develop a static general equilibrium model of monopolistic competition with heterogeneous firms, building on the canonical framework of Melitz, 2003. This model provides a tractable, closed-form mapping between AI, conceptualized as a cognitive General Purpose Technology (GPT), and the macroeconomic dynamics of entry, pricing, and aggregate welfare. While we present the core economic intuition and the main propositions below, the step-by-step mathematical derivation of the equilibrium is provided in Appendix C.

5.1 Environment and Equilibrium

Consider an economy populated by a representative household that supplies L units of labor inelastically. Labor is the sole factor of production, and we set the nominal wage as the numéraire ($W = 1$).[5] The household allocates its income over a continuum of available service varieties Ω, whose endogenous total mass is N. Preferences take the standard Constant

[5]By normalizing the nominal wage and assuming inelastic labor supply, we intentionally abstract from nominal labor market frictions and employment displacement dynamics. This abstraction allows us to cleanly isolate the transmission of the technological shift through product markets via firm entry and pricing.

Elasticity of Substitution (CES) form:

$$U = \left(\int_{i \in \Omega} q_i^{\frac{\varepsilon-1}{\varepsilon}} di \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{10}$$

where $\varepsilon > 1$ is the elasticity of substitution between varieties.

On the production side, there exists an unbounded pool of prospective entrants. To produce quantity q_i, a firm requires an overhead fixed cost of operation f and a constant marginal cost $c_i = 1/A_i$, both measured in labor units. Here, A_i denotes the firm's idiosyncratic labor productivity. Facing CES demand, a firm with productivity A_i maximizes profits by setting a constant markup over marginal cost, yielding an operating profit strictly increasing in its productivity: $\pi(A_i) = BA_i^{\varepsilon-1} - f$, where B is an endogenous aggregate demand shifter.

Firms do not know their productivity prior to entry. To discover A_i, an entrepreneur must pay an upfront sunk entry cost f_e (capturing initial R&D, legal setup, and cognitive overhead). Upon paying f_e, the firm draws A_i from a common Pareto distribution with cumulative density function $G(A) = 1 - (A_{min}/A)^k$ and shape parameter $k > \varepsilon - 1$. A firm will only operate if its operating profit covers the fixed overhead cost, defining a Zero-Cutoff Profit (ZCP) survival threshold A^*. Simultaneously, a Free Entry (FE) condition dictates that the ex-ante expected profit of taking a productivity draw must exactly equal the sunk entry cost f_e.

Solving this system of conditions under the Pareto distribution yields the pre-AI equilibrium mass of operating firms:

$$N_{pre} = \frac{L}{\varepsilon f} \left(\frac{k - (\varepsilon - 1)}{k} \right) \tag{11}$$

5.2 The Macroeconomic Effects of Generative AI Adoption

We model the advent of generative AI as a symmetric, sector-specific technological integration that augments cognitive labor efficiency by a factor of $(1 + \gamma)$, where $\gamma > 0$. Because labor is the sole input, this adoption proportionately scales down all labor requirements. Specifically, the effective productivity of any firm increases to $A_i(1 + \gamma)$, thereby lowering marginal costs. Symmetrically, both the administrative fixed cost of operation and the bureaucratic sunk cost of entry fall to $f/(1 + \gamma)$ and $f_e/(1 + \gamma)$, respectively.

Proposition 1 (The Firm Entry Effect): *The introduction of a labor-saving cognitive technology ($\gamma > 0$) induces a proportional, sustained increase in the mass of operating firms. [Formal proof provided in Appendix C].*

Intuition: Substituting the new, lower operational overhead cost $f_{post} = f/(1+\gamma)$ into the equilibrium firm mass equation yields:

$$N_{post} = \frac{L}{\varepsilon\left(\frac{f}{1+\gamma}\right)}\left(\frac{k-(\varepsilon-1)}{k}\right) = N_{pre} \cdot (1+\gamma) \tag{12}$$

This analytical result cleanly formalizes our empirical findings in Section 3: the technological reduction in administrative and entry barriers triggers a direct, structural influx of new firms ($\partial N/\partial\gamma > 0$).

Proposition 2 (The Disinflationary Effect): *The technological integration reduces the aggregate price index through a dual channel: direct marginal cost reductions and enhanced product variety. [Formal proof provided in Appendix C].*

Intuition: The aggregate price index in the post-adoption equilibrium incorporates both the expanded mass of firms N_{post} and the augmented effective productivity $\tilde{A}_{post} = \tilde{A}_{pre}(1+\gamma)$. Substituting these into the aggregate CES price index yields:

$$P_{post} = P_{pre} \cdot (1+\gamma)^{-\frac{\varepsilon}{\varepsilon-1}} \tag{13}$$

Because $\varepsilon > 1$, it follows strictly that $\partial P/\partial\gamma < 0$. The aggregate price level falls due to the pure marginal cost reduction ($(1+\gamma)^{-1}$) amplified by the love-for-variety effect ($(1+\gamma)^{-\frac{1}{\varepsilon-1}}$) generated by the new entrants. This mechanism formally corroborates our empirical findings regarding relative price suppression in Section 4.

Corollary (The Welfare Dividend): Utility maximization in this framework implies that aggregate real welfare is equivalent to total real labor earnings ($U = L/P$). Since the aggregate labor supply L is inelastic, welfare is entirely driven by the real wage ($1/P$). Consequently, the technological integration unambiguously increases aggregate real welfare through two compounding forces: an increase in the number of available varieties N (the standard "love for variety" effect) and the pass-through of cost reductions to consumer prices. This theoretical result highlights a substantial and structurally persistent consumer dividend.

6 Conclusion

The rapid commercialization of generative artificial intelligence has initiated one of the most profound technological transformations in modern economic history. While the initial wave of economic research has predominantly—and understandably—fixated on the disruptive consequences of cognitive automation for the labor market, this paper broadens the analytical lens. By constructing a unified empirical and theoretical framework, we demonstrate that generative AI acts as a potent catalyst for product-market dynamism. Our granular empirical analysis reveals that occupational exposure to cognitive automation triggers an immediate, structural wave of firm entry, substantially lowering the barriers for both independent entrepreneurs and complex employer ventures. Concurrently, our dynamic event-study design confirms that these firm-level efficiencies pass through to consumers, generating a persistent relative price deflation in highly exposed sectors.

Through the lens of our general equilibrium framework, these dual findings formalize a substantial and persistent consumer welfare dividend. This theoretical result provides a necessary macroeconomic counterweight to the prevailing pessimism surrounding artificial intelligence. While the transitional frictions of labor displacement and occupational restructuring are undeniably significant and can be expected to dominate near-term dynamics, the structural dividends of this General Purpose Technology should progressively unfold over a longer horizon. Ultimately, the initial firm-level cost reductions generated by cognitive automation act merely as the catalyst for a broader macroeconomic transformation, as these localized efficiencies are transmitted to the aggregate economy through heightened market competition, expanded product variety, and enhanced purchasing power.

Crucially, the empirical estimates presented in this paper constitute a conservative early evaluation. Economic history illustrates that previous sweeping technologies typically required decades of organizational restructuring and intangible capital accumulation to fully permeate the aggregate economy. While the diffusion of generative AI may proceed at a substantially accelerated pace, current enterprise adoption metrics indicate that the U.S. economy is still navigating the initial stages of this integration. Because we already identify transformative, statistically robust product-market responses at a relatively limited level of aggregate penetration, the magnitude of this structural dynamism is poised to compound significantly over the coming years.

These dynamics carry critical implications for both policy and future research. A fundamental value of this early analysis lies in anticipating macro-trends before they fully mature, enabling the design of forward-looking policies and institutional frameworks that foster this

technological transition. Specifically, for macroeconomic policymakers and central banks, our findings highlight generative AI not only as a powerful engine for business dynamism but also as a structural disinflationary anchor that must be accounted for in long-term inflation expectations. For the academic literature, this paper underscores the necessity of moving beyond partial-equilibrium labor models. Future research must seek to evaluate the race between the disruptive frictions of labor displacement and the compounding structural benefits of the consumer dividend documented here. Understanding how these opposing forces co-evolve across the technological diffusion curve will be essential to fully grasp the aggregate macroeconomic trajectory of the AI era.

References

Acemoglu, D. (2024). *The simple macroeconomics of AI*. Tech. rep. National Bureau of Economic Research.

Acemoglu, D. and P. Restrepo (2018). "Artificial intelligence, automation, and work". *The economics of artificial intelligence: An agenda*, pp. 197–236.

– (2020). "Robots and jobs: Evidence from US labor markets". *Journal of Political Economy* 128.6, pp. 2188–2244.

Autor, D. (2024). *AI and the expertise economy*. Working Paper 32140. National Bureau of Economic Research.

Autor, D. H. (2013). "The labor market impacts of technological change: From unmeasured quality change to meaningful economic growth". *Annals of Economics and Statistics*, pp. 4–9.

Autor, D. H., J. J. Donohue III, and S. J. Schwab (2006). "The costs of wrongful-discharge laws". *The Review of Economics and Statistics* 88.2, pp. 211–231.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024). *Artificial intelligence, firm growth, and product innovation*. Tech. rep. National Bureau of Economic Research.

Bayard, K., E. Dinlersoz, T. Dunne, J. Haltiwanger, J. Miranda, and J. Stevens (2018). *Early-stage business formation: An analysis of applications for employer identification numbers*. Tech. rep. National Bureau of Economic Research.

Brynjolfsson, E. (1993). "The productivity paradox of information technology". *Communications of the ACM* 36.12, pp. 66–77.

Brynjolfsson, E., D. Li, and L. R. Raymond (2023). *Generative AI at work*. Tech. rep. National Bureau of Economic Research.

Brynjolfsson, E., D. Rock, and C. Syverson (2021). "The productivity J-curve: How intangibles complement general purpose technologies". *American Economic Journal: Macroeconomics* 13.1, pp. 333–372.

David, P. A. (1990). "The dynamo and the computer: an historical perspective on the modern productivity paradox". *The American Economic Review* 80.2, pp. 355–361.

Decker, R., J. Haltiwanger, R. Jarmin, and J. Miranda (2014). "The role of entrepreneurship in US job creation and economic dynamism". *Journal of Economic Perspectives* 28.3, pp. 3–24.

Eloundou, T., S. Manning, P. Mishkin, and D. Rock (2023). "GPTs are GPTs: An early look at the labor market impact potential of large language models". *arXiv preprint arXiv:2303.10130*.

Federal Open Market Committee (2023). *Minutes of the Federal Open Market Committee, December 12–13, 2023*. Meeting Minutes. Board of Governors of the Federal Reserve System.

Federal Reserve Bank of New York (2023). *Global Supply Chain Pressure Index (GSCPI)*. `https://www.newyorkfed.org/research/policy/gscpi`. Data release and historical series, accessed March 2026.

Graetz, G. and G. Michaels (2018). "Robots at work". *The Review of Economics and Statistics* 100.5, pp. 753–768.

Griliches, Z. and J. A. Hausman (1986). "Errors in variables in panel data". *Journal of econometrics* 31.1, pp. 93–118.

Haltiwanger, J., R. S. Jarmin, and J. Miranda (2013). "Who creates jobs? Small versus large versus young". *The Review of Economics and Statistics* 95.2, pp. 347–361.

Hopenhayn, H. A. (1992). "Entry, exit, and firm dynamics in long run equilibrium". *Econometrica: Journal of the Econometric Society*, pp. 1127–1150.

Melitz, M. J. (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity". *Econometrica* 71.6, pp. 1695–1725.

Mokyr, J. (1990). *The lever of riches: Technological creativity and economic progress*. Oxford University Press.

Oliner, S. D., D. E. Sichel, and K. J. Stiroh (2007). "Explaining a productive decade". *Brookings Papers on Economic Activity* 2007.1, pp. 81–137.

Appendices

A Data Construction and Sources

A.1. Sectoral AI Exposure Construction

The construction of the sectoral AI exposure index relies on merging occupational scores from Eloundou et al., 2023 with the BLS OEWS May 2022 dataset. The original O*NET-SOC occupational codes from the OpenAI dataset (typically 10 digits, e.g., 11-1011.00) are truncated to their first 7 digits to perfectly match the detailed occupational classification used by the BLS. In cases where multiple 10-digit O*NET occupations collapse into a single 7-digit BLS occupation, we take the unweighted mean of the exposure scores before merging.

From the BLS OEWS dataset, we restrict the sample to detailed occupations to avoid double counting nested hierarchies. We drop any occupational cells where total employment or mean annual wages are suppressed by the BLS for confidentiality reasons. The wage bill for each occupation-industry cell is calculated as the product of total employment and the mean annual wage.

Table A.1 reports the summary statistics of the resulting AI exposure index at both the 2-digit and 3-digit NAICS aggregation levels. The mean exposure is approximately 0.35 across sectors, indicating that, on average, 35% of the sectoral wage bill is allocated to occupational tasks highly exposed to generative AI. The significant standard deviation (approximately 0.10) confirms the presence of substantial cross-sectional variation in AI susceptibility across the U.S. economy, providing the necessary identifying variation for our analysis.

A.2. Aggregation of Business Applications

Data on Business Applications were obtained from the Federal Reserve Economic Data (FRED) database, sourced originally from the U.S. Census Bureau's Business Formation Statistics. We download the seasonally adjusted monthly series for both Total Business Applications (`BABANAICS*SAUS`) and High-Propensity Business Applications (`BAHBANAICS*SAUS`).

The dataset natively provides 19 broad sectoral series. Consistent with Census Bureau definitions, the manufacturing sector (NAICS 31-33), retail trade (NAICS 44-45), and transportation and warehousing (NAICS 48-49) are pre-aggregated. We exclude the

Table A.1: Summary Statistics: Sectoral AI Exposure Index

	NAICS 2-Digit	NAICS 3-Digit
Observations	20	85
Mean	0.362	0.350
Std. Dev.	0.107	0.104
Min	0.174	0.172
25th Percentile	0.296	0.273
Median	0.342	0.336
75th Percentile	0.423	0.402
Max	0.541	0.583

Notes: The table reports unweighted cross-sectional summary statistics for the constructed Generative AI exposure index (β) at the 2-digit and 3-digit NAICS levels. The index represents the wage-bill-weighted share of occupational tasks exposed to AI within each sector.

public administration sector, resulting in a balanced panel of 19 distinct private sectors. Monthly flows are summed within each calendar year to construct the annual measure of entry ($BA_{j,t} = \sum_{m=1}^{12} BA_{j,m,t}$). The Low-Propensity Applications series is mechanically constructed at the annual level as $LPA_{j,t} = BA_{j,t} - HPA_{j,t}$.

A.3. Hierarchical Reconstruction of Sectoral Producer Prices

Of the 85 3-digit NAICS sectors present in our baseline AI exposure index, exact aggregate Producer Price Indices (PPI) are natively available in FRED for 48 sectors. For the remaining sectors, the BLS often publishes price indices only at finer levels of disaggregation (4-, 5-, or 6-digit NAICS) or restricts publication due to confidentiality and sampling limitations.

To prevent the systematic loss of these sectors, we implement a hierarchical reconstruction algorithm. For any missing 3-digit sector j, we identify the universe of its available sub-sectoral PPI series in the BLS master catalog. To avoid double-counting within nested hierarchies, we apply a strict survival filter: a sub-sectoral code is retained if and only if no broader parent code is available in the dataset.

Let Ω_j denote the set of structurally independent sub-sectors identified for the missing 3-digit sector j. We reconstruct the aggregate price index by computing the unweighted average of the monthly growth rates of the components in Ω_j:

$$\frac{\Delta P_{j,t}}{P_{j,t-1}} = \frac{1}{|\Omega_j|} \sum_{k \in \Omega_j} \frac{\Delta P_{k,t}}{P_{k,t-1}} \tag{A.1}$$

The reconstructed price index level $P_{j,t}$ is then recovered via cumulative multiplication. We favor an unweighted average over employment or value-added weighting schemes at this highly disaggregated level because dynamic revenue weights are rarely available at a monthly frequency for 6-digit NAICS sub-sectors, and static weights risk introducing mechanical measurement error into high-frequency price indices.

This procedure successfully recovers 17 additional 3-digit NAICS sectors, bringing our final analytical sample to 65 sectors. Table A.2 documents the exact composition of the reconstructed aggregates. The remaining 20 sectors (primarily non-market services, education, and specific public-adjacent sectors) are dropped from the panel due to a complete absence of reliable BLS producer price data.

Table A.2: Composition of Reconstructed 3-Digit NAICS Price Indices

3-Digit NAICS	**Constituent Sub-Sector Codes Used for Aggregation**
113	1133
236	236211, 236221, 236222, 236223, 236224, 236400, 236500
238	23811X, 23816X, 2381MR, 23821X, 23822X
455	4551, 4552
458	45811
459	45940
518	51821
522	5221
531	53112, 53113, 5312, 53131, 531320
532	5321, 532412
541	5411, 541211, 5413, 5416, 54181
561	5613, 56151, 561612, 56172
562	5621
621	6211, 6212, 6215, 6216, 62199
623	6231, 62321
713	7131, 71391, 71394
811	8113

Notes: This table lists the 17 3-digit NAICS sectors for which aggregate producer price indices were reconstructed. The right column details the exact BLS PCU sub-sector codes (Ω_j) that survived the hierarchical filter and were geometrically averaged to compute the 3-digit aggregate. Codes ending in 'X' or 'MR' follow standard BLS PCU nomenclature for specific industry groupings.

B Empirical Robustness

Table B.1: Robustness: Leave-One-Out and Cluster Exclusion

Excluded Sector(s)	**AI Exposure Coef. (β_1)**	**p-value**
Panel A: Leave-One-Out (Single Sector)		
11 (Agriculture)	0.512**	0.013
21 (Mining)	0.753***	0.008
22 (Utilities)	0.739***	0.007
23 (Construction)	0.704**	0.020
31-33 (Manufacturing)	0.728**	0.010
42 (Wholesale Trade)	0.747***	0.009
44-45 (Retail Trade)	0.748***	0.007
48-49 (Transportation)	0.704**	0.017
51 (Information)	0.703**	0.020
52 (Finance)	0.834***	0.004
53 (Real Estate)	0.753***	0.008
54 (Professional Services)	0.729**	0.016
55 (Management)	0.698**	0.021
56 (Admin & Support)	0.717**	0.010
61 (Educational Services)	0.733***	0.009
62 (Health Care)	0.737***	0.009
71 (Arts & Entertainment)	0.777***	0.006
72 (Accommodation/Food)	0.857***	0.003
81 (Other Services)	0.705**	0.012
Panel B: Dual Exclusion (Knowledge-Intensive Clusters)		
51 & 54	0.694**	0.041
52 & 55	0.811**	0.013
53 & 56	0.736***	0.010

Notes: The table reports the point estimate and the robust p-value for the generative AI exposure coefficient (β) from iterative re-estimations of our baseline model (Table 1, Column 1). Panel A systematically excludes one NAICS sector at a time. Panel B jointly excludes specific pairs of knowledge-intensive clusters. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2: Robustness: Alternative Trends and Direct Growth Rates

	Quad. Gap	Growth (2023-2025)		Growth Avg. (22/23-24/25)	
	(1) Quad. Trend	(2) Short CAGR	(3) Long CAGR	(4) Short CAGR	(5) Long CAGR
Generative AI Exposure (β)	0.557**	0.918**	0.956**	0.590***	0.610***
	(0.221)	(0.349)	(0.347)	(0.169)	(0.175)
Observations	19	19	19	19	19
R^2	0.417	0.601	0.578	0.585	0.565

Standard errors in parentheses

Controls include Labor Share and the respective historical pre-trend or CAGR.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Column (1) estimates the baseline gap using a quadratic pre-trend. Columns (2) and (3) estimate the direct log difference in business applications from 2023 to 2025. Columns (4) and (5) estimate the log difference between the 2022-2023 average and the 2024-2025 average. Controls include the Labor Share and the corresponding pre-2020 historical compound annual growth rate (CAGR). Heteroskedasticity-robust standard errors are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.1 Event Study

Table B.3 reports the full set of point estimates and standard errors for the dynamic event study presented in Section 4.2. The results confirm the absence of statistically significant pre-trends prior to the 2019 base year.

Furthermore, Figure B.1 presents an iteration of our baseline event study that extends the sample to include the first available observation of 2026 (January). While our baseline specification conservatively truncates the sample at 2025 to ensure all temporal periods represent full-year unweighted averages, incorporating the January 2026 data point offers suggestive evidence regarding the ongoing trajectory of the technological effect. The continued downward trajectory of the point estimate provides suggestive evidence that the disinflationary effect of generative AI is actively compounding.

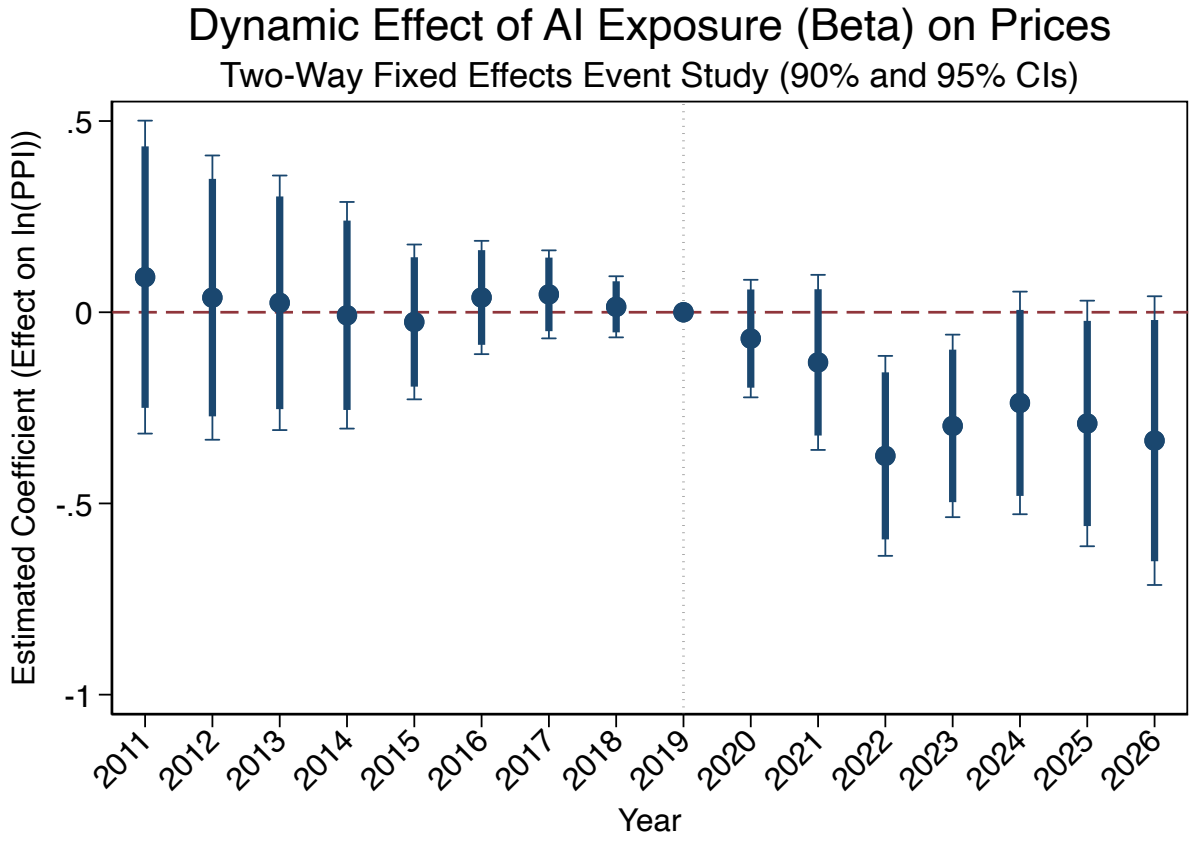


Figure B.1: Dynamic Event Study: Extended Sample (Including Jan 2026)

Table B.3: Dynamic Event Study: Full Estimation Results

	(1) ln(PPI)
$\beta \times \mathbb{I}\{2011\}$	0.092
	(0.205)
$\beta \times \mathbb{I}\{2012\}$	0.038
	(0.186)
$\beta \times \mathbb{I}\{2013\}$	0.025
	(0.167)
$\beta \times \mathbb{I}\{2014\}$	-0.008
	(0.148)
$\beta \times \mathbb{I}\{2015\}$	-0.025
	(0.101)
$\beta \times \mathbb{I}\{2016\}$	0.039
	(0.074)
$\beta \times \mathbb{I}\{2017\}$	0.047
	(0.058)
$\beta \times \mathbb{I}\{2018\}$	0.014
	(0.040)
$\beta \times \mathbb{I}\{2019\}$ (Base)	0.000
	(.)
$\beta \times \mathbb{I}\{2020\}$	-0.069
	(0.077)
$\beta \times \mathbb{I}\{2021\}$	-0.131
	(0.115)
$\beta \times \mathbb{I}\{2022\}$	-0.375***
	(0.131)
$\beta \times \mathbb{I}\{2023\}$	-0.297**
	(0.119)
$\beta \times \mathbb{I}\{2024\}$	-0.237
	(0.145)
$\beta \times \mathbb{I}\{2025\}$	-0.291*
	(0.159)
Sector FE	Yes
Year FE	Yes
Number of groups	65
Observations	923
Within R^2	0.703

Notes: The table reports the coefficients from the continuous difference-in-differences specification (Equation 9). The dependent variable is the log Producer Price Index ($\ln(PPI)$). The year 2019 is the omitted base category. Standard errors, clustered at the NAICS industry level, are reported in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Theoretical Appendix: Model Derivations

This appendix provides the formal mathematical derivations for the theoretical framework discussed in Section 5. We proceed by solving the firm's optimization problem, defining the equilibrium conditions under a Pareto productivity distribution, and formally proving the comparative statics associated with the generative AI technological adoption.

C.1 Firm Pricing and Relative Profits

The representative household allocates aggregate expenditure R across a continuum of varieties Ω. Solving the standard CES utility maximization problem yields the demand function for any individual variety i:

$$q_i = \frac{R}{P}\left(\frac{p_i}{P}\right)^{-\varepsilon} \tag{C.1}$$

where $P = \left(\int_{i\in\Omega} p_i^{1-\varepsilon} di\right)^{\frac{1}{1-\varepsilon}}$ is the aggregate price index, and $\varepsilon > 1$ is the elasticity of substitution.

Firms produce using only labor, which is supplied inelastically by households (L). Normalizing the wage rate to unity ($W = 1$), the total cost function for a firm with idiosyncratic productivity A_i is $C(q_i) = f + q_i/A_i$. Facing downward-sloping CES demand, the firm maximizes operating profits $\pi_i = p_i q_i - q_i/A_i - f$. The first-order condition yields the standard constant markup pricing rule:

$$p(A_i) = \left(\frac{\varepsilon}{\varepsilon - 1}\right)\frac{1}{A_i} \tag{C.2}$$

Substituting this pricing rule and the demand function back into the profit equation, we can express a firm's operating profit strictly as a function of its productivity draw:

$$\pi(A_i) = B \cdot A_i^{\varepsilon-1} - f \tag{C.3}$$

where $B \equiv \frac{R}{\varepsilon}\left(P\frac{\varepsilon-1}{\varepsilon}\right)^{\varepsilon-1}$ acts as an endogenous aggregate demand shifter that is common to all firms.

C.2 The Zero-Cutoff Profit and Free Entry Conditions

Firms will only produce if their operating profits are non-negative. This establishes the Zero-Cutoff Profit (ZCP) condition, which defines the survival threshold A^*, representing

the productivity of the marginal firm that makes exactly zero profits:

$$\pi(A^*) = 0 \quad \Longrightarrow \quad B \cdot (A^*)^{\varepsilon-1} = f \tag{C.4}$$

Using the ZCP condition, we can substitute out the aggregate demand shifter B. This allows us to express the operating profit of any surviving firm ($A_i \geq A^*$) relative to the marginal firm:

$$\pi(A_i) = f \left[\left(\frac{A_i}{A^*} \right)^{\varepsilon-1} - 1 \right] \tag{C.5}$$

Prior to entry, firms face uncertainty regarding their productivity. To take a productivity draw from the cumulative distribution function $G(A)$, an entrant must pay a sunk cost f_e. In a free-entry equilibrium, the ex-ante expected profit of entering the market must be driven to exactly equal this sunk cost. This defines the Free Entry (FE) condition:

$$\int_{A^*}^{\infty} \pi(A) dG(A) = f_e \tag{C.6}$$

Substituting the relative profit function (Equation C.5) into the FE condition (Equation C.6) yields the fundamental equilibrium relationship that determines the survival cutoff:

$$f \int_{A^*}^{\infty} \left[\left(\frac{A}{A^*} \right)^{\varepsilon-1} - 1 \right] dG(A) = f_e \tag{C.7}$$

C.3 Equilibrium under a Pareto Distribution

To obtain closed-form analytical solutions for the macroeconomic variables of interest, we follow standard convention and parameterize the ex-ante productivity draws using a Pareto distribution. The cumulative density function is given by:

$$G(A) = 1 - \left(\frac{A_{min}}{A} \right)^k \quad \text{for } A \geq A_{min} \tag{C.8}$$

where $A_{min} > 0$ is the lower bound of the support and $k > \varepsilon - 1$ is the shape parameter governing the dispersion of productivities (a higher k implies lower variance).

Substituting the Pareto density $dG(A) = kA_{min}^k A^{-k-1} dA$ into the Free Entry condition

(Equation C.7) and evaluating the integral over the domain $[A^*, \infty)$ yields:

$$f\left[\frac{\varepsilon-1}{k-(\varepsilon-1)}\right]\left(\frac{A_{min}}{A^*}\right)^k = f_e \tag{C.9}$$

Rearranging for A^* provides the explicit solution for the equilibrium survival threshold:

$$A^* = A_{min}\left[\frac{f}{f_e}\left(\frac{\varepsilon-1}{k-(\varepsilon-1)}\right)\right]^{\frac{1}{k}} \tag{C.10}$$

To determine the aggregate mass of operating firms (N), we impose the labor market clearing condition. In a free-entry equilibrium, aggregate expected corporate profits net of entry costs are zero. Consequently, total household expenditure must exactly equal total labor income, such that $R = L$.

The average operating profit of surviving firms, $\bar{\pi}$, can be derived by integrating the relative profit function (Equation C.5) over the truncated Pareto distribution for active firms ($A \geq A^*$):

$$\bar{\pi} = f\left[\frac{k}{k-(\varepsilon-1)} - 1\right] = f\left(\frac{\varepsilon-1}{k-(\varepsilon-1)}\right) \tag{C.11}$$

From the firm's profit maximization problem, average firm revenue is proportional to average operating profits and fixed costs: $\bar{r} = \varepsilon(\bar{\pi} + f)$. Substituting the expression for $\bar{\pi}$ derived above, the average firm revenue simplifies to a constant determined by structural parameters:

$$\bar{r} = \varepsilon f\left(\frac{k}{k-(\varepsilon-1)}\right) \tag{C.12}$$

Finally, the equilibrium mass of operating firms in the pre-AI economy is determined by dividing total market revenue by the average firm revenue ($N_{pre} = R/\bar{r}$). Recalling that $R = L$, we obtain:

$$N_{pre} = \frac{L}{\varepsilon f}\left(\frac{k-(\varepsilon-1)}{k}\right) \tag{C.13}$$

C.4 Generative AI Adoption

We model the integration of generative AI as the adoption of a cognitive General Purpose Technology that symmetrically scales the effective labor productivity of all firms by a factor of $(1+\gamma)$, where $\gamma > 0$. Because labor is the sole factor of production, this technological shift proportionately reduces all labor requirements.

For any firm i, effective productivity becomes $A_i(1+\gamma)$, naturally reducing the marginal cost of production to $c_i = 1/[A_i(1+\gamma)]$. Concurrently, the cognitive labor-saving nature of the technology reduces the administrative fixed cost of operation to $f_{post} = f/(1+\gamma)$, and the bureaucratic and research-driven sunk cost of entry to $f_{e,post} = f_e/(1+\gamma)$.

Proof of Proposition 1 (The Firm Entry Effect):
To determine the new survival threshold, we evaluate the equilibrium cutoff condition (Equation 10) under the post-adoption cost structure. Because the technological integration reduces operational and entry costs symmetrically, their relative ratio remains constant:

$$\frac{f_{post}}{f_{e,post}} = \frac{f/(1+\gamma)}{f_e/(1+\gamma)} = \frac{f}{f_e} \tag{C.14}$$

Consequently, the baseline survival threshold is structurally unaffected by the technological adoption ($A^*_{post} = A^*_{pre}$). However, substituting the new operational fixed cost f_{post} into the equilibrium firm mass equation (Equation 13) yields:

$$N_{post} = \frac{L}{\varepsilon\left(\frac{f}{1+\gamma}\right)}\left(\frac{k-(\varepsilon-1)}{k}\right) = N_{pre}\cdot(1+\gamma) \tag{C.15}$$

Since $\gamma > 0$, it strictly follows that $N_{post} > N_{pre}$. The permanent reduction in fixed operational barriers naturally supports a proportionally larger mass of equilibrium varieties. ■

Proof of Proposition 2 (The Disinflationary Effect):
Following Melitz, 2003, the aggregate CES price index can be expressed as a function of the equilibrium firm mass and the price charged by a firm with the weighted average productivity $\tilde{A}$:

$$P = N^{-\frac{1}{\varepsilon-1}}p(\tilde{A}) = N^{-\frac{1}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)\frac{1}{\tilde{A}} \tag{C.16}$$

Because the survival threshold A^* is unchanged (as proven in Proposition 1), the baseline average productivity of the surviving firm distribution remains $\tilde{A}_{pre}$. However, the post-adoption effective productivity of every firm is scaled by $(1+\gamma)$, such that $\tilde{A}_{post} = \tilde{A}_{pre}(1+\gamma)$.

Substituting the post-adoption firm mass (N_{post}) and effective average productivity ($\tilde{A}_{post}$) into the aggregate price index yields:

$$P_{post} = [N_{pre}(1+\gamma)]^{-\frac{1}{\varepsilon-1}}\left(\frac{\varepsilon}{\varepsilon-1}\right)\frac{1}{\tilde{A}_{pre}(1+\gamma)} \tag{C.17}$$

Rearranging and collecting the γ terms, we obtain:

$$P_{post} = \underbrace{\left[N_{pre}^{-\frac{1}{\varepsilon-1}} \left(\frac{\varepsilon}{\varepsilon-1} \right) \frac{1}{\tilde{A}_{pre}} \right]}_{P_{pre}} \cdot (1+\gamma)^{-\frac{1}{\varepsilon-1}} \cdot (1+\gamma)^{-1} \tag{C.18}$$

$$P_{post} = P_{pre} \cdot (1+\gamma)^{-\frac{\varepsilon}{\varepsilon-1}} \tag{C.19}$$

Since $\varepsilon > 1$ and $\gamma > 0$, the exponent is strictly negative, ensuring $P_{post} < P_{pre}$. Equation (C.18) formally isolates the dual disinflationary channel: prices fall due to the direct marginal cost reduction ($(1+\gamma)^{-1}$) multiplied by the love-for-variety effect generated by the influx of new entrants ($(1+\gamma)^{-\frac{1}{\varepsilon-1}}$).

Finally, because nominal labor income is the numéraire ($W = L = R$), aggregate real welfare is simply $U = L/P$. Since L is constant and $P_{post} < P_{pre}$, real welfare unambiguously rises. ■