Entrepreneurship, Experimentation, and Generative AI: Evidence from ChatGPT's Launch*

Andrea Contigiani (Ohio State University)
Gavin Wang (Purdue University)
Jiayi Bao (Texas A&M University)

July 28, 2026

Abstract

We study how generative AI (GenAI) reshapes the process of experimentation in early-stage ventures. We propose three mechanisms: an efficiency channel, where GenAI lowers the cost of each experiment; a precision channel, where GenAI increases the learning extracted per experiment; and a substitution channel, where GenAI replaces experimentation with non-experimental learning. We build a firm-month panel of 26,873 U.S. software ventures, combining Crunchbase and Revelio Labs with several additional data sources. We exploit the November 2022 launch of ChatGPT as a quasi-natural experiment and quantify exposure to the shock based on ventures' top management teams' background. After the launch, AI-ready ventures reduce the share of their workforce devoted to experimentation by roughly 13%, a decline concentrated in prototyping roles. The engineering share of the team falls in parallel. Adoption of experimentation tooling does not accelerate, while commerce infrastructure appears to rise. Product introduction, proxied by trademark filings, shows no detectable change. Financing shows no launch-timed break, while the hazard of exit rises by roughly 40% net of these ventures' elevated baseline exit propensity. Overall, the evidence is consistent with efficiency and substitution, while there is no support for precision.

*This manuscript is under active development and continues to undergo substantial revision. The latest version is available here.

1 Introduction

The advent of generative AI (GenAI), exemplified by large language models such as ChatGPT, is transforming how ventures are conceived, tested, and built: tools initially viewed as productivity aids for individual founders now permeate core entrepreneurial processes, from ideation and product design to customer interaction and fundraising. Yet despite the rapid diffusion of GenAI across the venture ecosystem, basic questions about how it will reshape the way startups are built remain unresolved: whether AI-native startups will need fewer or more employees, how team structures will evolve, and whether the cost of building and scaling a new venture will rise or fall (Chen, 2025). Addressing these questions requires moving beyond the observation that GenAI makes entrepreneurs faster and cheaper, to examine how it reshapes the ways in which they search, learn, and make strategic decisions.

A central activity in early-stage ventures is *experimentation*. A large literature examines the role of experimentation in driving venture performance and growth (Camuffo et al., 2020; Contigiani & Levinthal, 2019; Koning et al., 2022). In many settings, this approach is formalized through the "Lean Startup" methodology (Blank, 2013; Ries, 2011, 2017), which emphasizes iterative cycles of hypothesis formulation, MVP testing with real customers, and rapid adaptation to feedback. Experimentation is a primary mechanism through which entrepreneurs resolve uncertainty about whether their product, their market, and their business model are viable.

We explore how the emergence of GenAI affects ventures' reliance on experimentation. On the one hand, cheaper and more capable AI tools may lower the cost of running experiments at scale. On the other hand, as GenAI becomes more capable of generating and processing information, the informational value of costly field experiments may diminish. In this paper, we ask how the availability of GenAI changes the human capital early-stage ventures devote to experimentation.

We propose three mechanisms through which GenAI may alter the demand for experimentation-related human capital. The *efficiency channel* views GenAI as a complement that lowers the cost of each experiment: by generating prototypes, marketing materials, and analytical scripts more cheaply, GenAI reduces the human effort required per unit of experimental learning. The *precision channel* operates on the output side: GenAI increases the learning extracted from each experiment through automated transcription, behavioral data synthesis, and AI-assisted insight generation, potentially reducing the number of experiments required. The *substitution channel* posits that GenAI crowds out traditional experimentation by raising the returns to

"non-experimental learning": synthetic customer feedback, scenario analysis, and expert-like advice let firms learn "on the screen" rather than "in the field." All three channels may reduce the share of the workforce allocated to experimentation, but the efficiency and precision channels also admit the opposite prediction: cheaper and more informative experiments may induce ventures to run more of them. Whether GenAI complements or substitutes for experimental learning is therefore difficult to predict ex ante.

To examine this question, we assemble a firm-month panel of 26,873 U.S. software ventures founded between 2018 and 2021, following them from founding through December 2025 for approximately 2.1 million firm-month observations. We exploit the November 2022 public release of ChatGPT as a sudden, economy-wide technological shock (Bao et al., 2025) in a difference-in-differences (DiD) design. Because the shock affects all ventures in principle, our empirical strategy requires identifying those that were more intensely exposed. We classify a venture as *AI-ready* if at least one top management team (TMT) member possessed deep learning (DL) or machine learning (ML) expertise prior to ChatGPT's launch. Drawing on the theory of absorptive capacity (Cohen & Levinthal, 1990), we argue that such ventures were better positioned to recognize, assimilate, and apply GenAI tools once they became widely accessible. We measure experimentation using job titles and descriptions from resume data (Wang & Wu, 2026). Using data from Revelio Labs, we identify positions involving experimentation-related tasks and distinguish three empirically distinct types: customer input gathering (surveys, user research), prototyping (MVP building, piloting, concept validation), and controlled testing (A/B tests and formal experiments).

Following the launch, the share of the workforce that AI-ready ventures allocate to experimentation declines by roughly 13% relative to its baseline. The effect is concentrated in prototyping-oriented roles and is accompanied by a parallel decline in the Engineering share of the team. The scaling share is unaffected, indicating a response specific to experimentation rather than a lifecycle shift. Downstream, website-technology data show no post-launch acceleration in experimentation tooling, which weighs against an expansionary response under either the efficiency or precision channel. Commerce infrastructure rises, and trademark-measured product introduction is unaffected in the observable window. The one downstream margin that moves is exit: after the launch, AI-ready ventures' hazard of acquisition or IPO rises by roughly 40% relative to their own elevated baseline. Taken together, the evidence finds no support for the precision channel but is consistent with fixed-quantity efficiency and with substitution.

This study makes three contributions. First, it provides one of the first large-scale firm-level analyses of how GenAI reshapes entrepreneurial search and learning. Whereas existing work has primarily examined

how ChatGPT spurred new firm creation or how AI tools affect individual labor outcomes, we document how a general-purpose technological shock restructures learning inside existing ventures: the workforce share devoted to experimentation falls, concentrated in the prototyping activities that anchor lean-startup learning and that GenAI's generative capabilities can perform. Second, we contribute to absorptive capacity theory and research on technological change and human capital by documenting how pre-existing DL/ML expertise among founders conditions firms' responses to a new general-purpose technology, shaping both the scale and composition of the capabilities they build in response. Third, the evidence speaks to the AI-and-jobs debate: inside firms, the margin we identify causally is compositional—GenAI hits specific task content while employment descriptively grows—and absorptive capacity determines where that recomposition lands first.

The remainder of the paper proceeds as follows. Section 2 develops the theoretical framework and the three mechanisms. Section 3 describes the empirical setting, including the emergence of ChatGPT and our approach to measuring differential exposure. Section 4 presents the data, sample construction, and key variables. Section 5 validates the treatment and lays out the empirical strategy. Section 6 reports the analysis: the main results on experimentation-related human capital and team composition, followed by website-technology adoption, product introduction, and financing and exit. Section 7 concludes by discussing key findings, contributions to research, and limitations.

2 Theory

We propose a simple theoretical framework linking the emergence of generative AI to ventures' demand for experimentation and, through it, for experimentation-related human capital. We first describe entrepreneurial experimentation as a Bayesian learning process, then propose three channels through which GenAI may reshape it.

2.1 Experimentation and Generative AI

Early-stage ventures learn by running experiments in their environment. As this practice gained widespread traction among practitioners in the 2010s (Blank, 2013; Ries, 2011, 2017), a growing body of research has examined the role of experimental learning in new venture development (Camuffo et al., 2020, 2024; Chavda et al., 2024; Contigiani & Levinthal, 2019; Eisenhardt & Bingham, 2017; Kerr et al., 2014). Against this

backdrop, we ask whether the nature of entrepreneurial experimentation is changing in the age of GenAI.

An essential premise of experimental learning is that there exists a target state of product-market fit, which the entrepreneur approaches through iterative cycles of prototyping, testing with targeted users, and incorporating the resulting feedback (Contigiani & Levinthal, 2019). Each cycle generates information that refines the entrepreneurial idea, progressively reducing uncertainty about whether the product, the market, and the business model are viable.

The process of experimentation is often described through a Bayesian lens (Agrawal et al., 2025; Contigiani et al., 2026). The founder holds a prior belief about the market—about whether the product, the customer segment, and the business model are viable. An experiment is the purchase of a signal, produced at cost K: the venture chooses to generate information about the market rather than act on beliefs alone. The founder then updates: the posterior formed from prior and signal becomes the prior of the next cycle, and the venture decides whether to persist, pivot, or abandon.

Practically, producing the signal takes three steps. In the *design* step, the venture constructs a test capable of generating evidence on the belief in question—building a minimum viable product, designing a customer interview protocol, or specifying the conditions of an A/B test. In the *execution* step, it deploys the test in the field, recruiting participants, collecting behavioral data, and recording customer responses. In the *measurement* step, it processes and analyzes the resulting data, extracting structured signal from raw observations.

A single experiment is thus one update: prior, signal (design, execution, measurement), posterior. Between birth and product-market fit, the venture runs some number N of experiments, each at cost K. N is not fixed but chosen: the venture continues until the expected value of a further signal no longer justifies its cost.

We propose three mechanisms through which GenAI may alter ventures' demand for experimentation. The *efficiency channel* operates at the design step: by reducing the cost of building prototypes, interfaces, and marketing assets, GenAI lowers K, though cheaper experiments may equally induce ventures to run more of them, leaving the effect on N ambiguous. The *precision channel* operates at the measurement step: by enabling richer extraction of signal from experimental data, GenAI increases the learning yield per cycle. Fewer experiments may be needed to reach a given posterior, but higher-quality information may raise the marginal value of further testing, again leaving the effect on N ambiguous. The *substitution channel* operates through a different logic: rather than making field experiments cheaper or more informative, GenAI raises the

returns to *non-experimental* learning by providing low-cost synthetic customer feedback, scenario analysis, and expert-like advice, allowing ventures to learn "on the screen" rather than in the field. Unlike the efficiency and precision channels, the substitution channel unambiguously predicts a reduction in the number of field experiments N.

2.2 The Efficiency Channel

The first mechanism, the *efficiency channel*, operates at the design step: GenAI lowers the cost K of building and deploying each experiment without changing the value of experimentation itself. The key insight is that a large share of K is consumed not by the act of learning, but by the act of constructing the vehicle for learning: writing the code for a prototype, designing the interface, drafting the copy, meeting regulatory requirements. GenAI compresses each of these input costs substantially.

The compression operates across each input to experiment construction. On the build margin, AI coding assistants scaffold a working prototype from a natural-language description: a founder can stand up a landing page, a clickable front-end, or a functioning demo in hours of assisted coding rather than weeks of engineering time. On the content margin, generative tools draft the assets an experiment consumes (ad copy variants for a smoke test, mock product imagery, onboarding email sequences) at near-zero marginal cost. On the preparation margin, AI-powered research tools compress the overhead that precedes fieldwork: drafting interview guides and screener surveys, generating test scripts, and summarizing the regulatory and platform requirements a pilot must satisfy. In each case the experiment itself is unchanged: the same artifact is put in front of the same real users, but the cost of constructing its vehicle falls.

The implication for the demand for experimentation is theoretically ambiguous. When K falls, standard economic reasoning predicts an increase in the quantity of signals purchased, a Jevons paradox in which efficiency gains stimulate rather than dampen experimental activity: a founder who previously could afford one MVP test per month might now run five. Alternatively, the venture may hold N fixed, reaching the same posterior along the same path at lower cost. Whether N rises or holds is ultimately an empirical question.

2.3 The Precision Channel

The second mechanism, the *precision channel*, operates at the measurement step: its defining feature is an increase in the learning yield of each experiment, the informativeness of the signal a given cycle produces.

The value of an experiment depends not only on what data it generates, but on how much structured signal can be extracted from that data. Raw experimental output (interview transcripts, behavioral logs, click-tracking data, usability recordings) is noisy, voluminous, and difficult to synthesize; historically, extracting actionable insight required substantial human effort. GenAI compresses this analytical burden, and because measurement is one step of signal production, the same compression may also reduce the measurement component of the cost K. What distinguishes precision from efficiency is therefore where it operates, the measurement step and the informativeness of the signal, rather than K being literally fixed.

The precision gain operates wherever raw experimental output awaits analysis. Interview pipelines that once required a researcher to transcribe, code, and synthesize by hand can now be processed at scale: an LLM digests dozens of hours of customer conversations and returns structured themes, recurring objections, and ranked feature requests. Usability evidence that once demanded dedicated review (session recordings, click paths, drop-off points) can be summarized into a prioritized list of friction points. And live tests can be read in something closer to real time, with natural-language querying of funnel and cohort data replacing bespoke analyst work between iterations. In each case the venture runs the same experiment on the same users but extracts more structured signal from it. The learning yield of a given cycle rises.

The implication for the demand for experimentation is again ambiguous. A more informative signal moves the posterior further per experiment: fewer experiments may be needed to reach the learning threshold at which the venture commits, so N may fall. But the Jevons logic applies here too: higher-quality information from each experiment raises the marginal value of running the next one—better measurement reveals new questions, uncovers new customer segments, or opens new product directions. A founder who previously stopped after three customer interviews because the signal was too noisy might now run ten. The direction of the effect on N is, as with the efficiency channel, an empirical matter.

2.4 The Substitution Channel

The third mechanism, the *substitution channel*, replaces experimentation with non-experimental signals. GenAI can replace the design, execution, and measurement steps entirely with screen-based alternatives: rather than building a prototype, deploying it with real users, and analyzing their responses, the AI-ready venture can simulate all three steps without leaving the screen. In Bayesian terms, the founder continues to update, but purchases the signal synthetically rather than from the field.

What substitution displaces is the prototype-test cycle *as a bundle*: design an artifact, deploy it,

measure the response. Each step has a screen-based counterpart. At the design step, GenAI generates and stress-tests product concepts before any artifact exists: a founder can describe a feature and receive variants, objections, and edge cases in minutes, without writing code. At the execution step, persona platforms simulate deployment: AI personas stand in for target customers, reacting to concepts, copy, and pricing as a segment plausibly would, and simulated interviews or focus groups surface responses without recruiting a single participant. At the measurement step, the model converts those synthetic responses into structured directional feedback of the kind a field test would have produced. The signal a venture would have purchased by building, piloting, and measuring a prototype can instead be purchased on the screen. This substitution logic is consistent with recent arguments that AI is creating an "era of abundant expertise," in which firms can access specialized knowledge through AI tools rather than building it internally (Yerramilli-Rao et al., 2025).

The implication for the demand for experimentation is unambiguous: field experiments are displaced by synthetic signals, so the number of experiments N the venture runs between birth and product-market fit falls. For ventures well-positioned to use GenAI, which we term *AI-ready*, this implies a reduced need to run the build-deploy-measure loop in the field. We note, however, that synthetic learning tools have recognized limitations: products that depend heavily on direct human experience (sensory feedback, emotional response, social interaction) cannot be adequately simulated, and AI-generated personas risk creating an echo chamber that amplifies rather than tests the founder's existing assumptions. The concentration prediction below therefore rests on an explicit fidelity assumption: synthetic signals substitute most readily for artifact-based prototyping and least readily for direct human elicitation and behavioral controlled testing, for exactly these reasons. Whether AI-ready ventures respect these limitations or substitute field learning prematurely is a question our data cannot directly resolve, but one we return to in the discussion.

2.5 Empirical Predictions

The channels above concern the demand for experimentation: the number of experiments N a venture runs and the cost K of producing each signal. Next, we discuss the link between the demand for experimentation and the need for human capital allocated to experimentation. We then derive what each channel predicts for the outcomes we observe.

The mapping is direct: since labor is the dominant component of K in early-stage ventures, the workforce a venture devotes to experimentation is, to a first approximation, the product $N \times K$, under the assumption that experimentation positions scale with the labor input to experiments. A channel can therefore

reduce the experimentation workforce along two distinct margins, by reducing the cost of each experiment (K) or by reducing the number of experiments (N), and the channels differ on exactly this point.

If the *efficiency channel* dominates, the effect on the experimentation workforce is ambiguous, because it depends on the response of N to the fall in K. Under the expansionary reading, cheaper signals induce more of them: the savings per experiment are partially or fully offset, and ventures should simultaneously expand complementary experimentation infrastructure (A/B testing tools, analytics platforms), invest in additional technology stack components, and shift incremental labor toward commercialization functions that exploit the gains from cheaper iteration. Under the fixed-quantity reading, N holds and the experimentation workforce falls with no complementary tooling buildout, a footprint that is observationally equivalent, in workforce and tooling data, to the substitution channel. Note that under either reading the efficiency channel's labor compression falls on design-step labor (prototype-building itself), so a decline concentrated in prototyping roles is consistent with fixed-quantity efficiency as well; the concentration pattern does not by itself distinguish it from substitution. Evidence on experimentation infrastructure therefore discriminates the expansionary reading from the other two, but cannot separate fixed-quantity efficiency from substitution.

If the *precision channel* dominates, the effect on the experimentation workforce is likewise ambiguous, tracking the response of N, with K unchanged or mildly reduced through its measurement component. If sharper signals reduce the number of cycles needed (each moves the posterior further, so fewer are needed), the experimentation workforce falls; if instead sharper signals raise the marginal value of further testing, the workforce rises with N. Because precision operates at the measurement step, its distinctive footprint—whatever the sign of the workforce response—is concentrated in measurement-intensive labor (controlled-testing and analytics roles), under the assumption that measurement labor scales with the number of experimental cycles, a prediction we deploy against the channel below. Under either response, moreover, we deploy an auxiliary assumption: if improved learning quality manifests in observable product-market outcomes within our window, the precision channel predicts stronger product introduction and financing. Sharper signals could instead induce earlier abandonment or quality improvements our proxies miss, so this prediction is assumption-dependent, a qualification we carry to the discussion.

If the *substitution channel* dominates, AI-ready ventures will reduce their experimentation human capital unambiguously: the cost K of the field experiments that remain is unchanged, but their number N falls as synthetic signals displace field signals, so the workforce $N \times K$ declines. Because tool adoption is sticky on the extensive margin, the accompanying tooling prediction is flat-to-mildly-declining experimentation

infrastructure (no complementary buildout) rather than sharp disinvestment. The concentration of the effect follows from the mechanism under the fidelity assumption stated above. What synthetic simulation reproduces is the informational output of the artifact-based prototyping loop (design an artifact, deploy it, measure the response), so substitution most readily displaces inductive, prototyping-oriented experimentation. Direct human elicitation retains distinctive value where products depend on real human experience (the human-experience and echo-chamber limits noted above), and deductive, hypothesis-driven controlled testing requires real behavioral data that simulation cannot supply. The predicted decline is therefore concentrated in prototyping-oriented experimentation, with customer input gathering and controlled testing comparatively spared (we label these three types Exp 1–3 when Section 4 defines the taxonomy).

Two features of this prediction structure shape what our data can and cannot identify. First, all three channels are consistent with a decline in the experimentation workforce: the substitution channel predicts it unconditionally, while the efficiency and precision channels predict it whenever the fall in K (efficiency) or in N (precision) outweighs any offsetting expansion in experimental activity. The workforce margin thus identifies that $N \times K$ fell, not whether N or K drove it. Second, the channels separate on their *downstream* predictions, which we test in Sections 6.3–6.5. The precision channel carries the one distinctive positive prediction, better product-market outcomes from better learning, so evidence on product introduction and financing bears on it directly. The efficiency and substitution channels, by contrast, leave the same footprint in workforce and tooling data: whether ventures run the same experiments at lower K or replace field experiments with synthetic signals, both imply a falling demand for experimentation human capital with no complementary tooling buildout. Separating those two requires observing experimental activity itself, which our data do not capture, a limitation we return to in Section 7. Table 1 summarizes the mechanisms and their predictions.

3 Setting

We describe the setting for our study. We begin by situating the launch of ChatGPT within the broader history of AI development, documenting the technical trajectory from the Transformer architecture to the GPT model family. We then discuss the validity of this event as a credible quasi-experimental shock and illustrate how we quantify exposure to the shock.

Table 1: Summary of Theoretical Mechanisms

	Efficiency	**Precision**	**Substitution**
Experimental Step	Design	Measurement	Design, Execution, Measurement
Mechanism	Reduces cost per signal	Increases learning per signal	Replaces experimental signals with non-experimental signals
Cost per Experiment (K)	Decreases	Unchanged or decreases (measurement component)	Unchanged
Number of Experiments (N)	Unchanged or Increases	Decreases or Increases	Decreases
Experimentation Workforce ($N \times K$)	Ambiguous	Ambiguous	Decreases

Notes: Summary of the channels through which GenAI affects the demand for experimentation and, through it, experimentation-related human capital. Each experiment produces a signal at cost K; the venture runs N experiments between birth and product-market fit. The efficiency channel operates at the design step; the precision channel at the measurement step; the substitution channel replaces the design, execution, and measurement steps with screen-based synthetic alternatives. The workforce row combines each channel's effect on K and on N.

3.1 The Emergence of Generative AI

Prior to late 2022, AI had been widely expected to transform the economy, yet the productivity statistics told a more modest story. Brynjolfsson et al. (2017) documented what they termed an "AI productivity paradox": despite substantial investment and high expectations, broad economic impact remained elusive. AI adoption was concentrated among large, sophisticated firms with access to specialized talent, and prior breakthroughs such as DeepMind's AlphaGo had captured public attention without meaningfully diffusing to mainstream economic activity (Babina et al., 2024). As of 2021, generative AI remained, for most of the economy, an unfulfilled promise.

The technical foundations for what would become ChatGPT were nevertheless being laid during this period. Vaswani et al. (2017) introduced the Transformer architecture, whose self-attention mechanism achieved state-of-the-art performance while enabling far more efficient scaling: the architecture processes sequences in parallel rather than token by token, relaxing the training bottlenecks of recurrent networks. OpenAI built directly on this work with GPT-1, which established the "pre-train on generic text, then adapt" recipe (Radford et al., 2018): a single model trained on unlabeled text and then fine-tuned cheaply for individual tasks, demonstrating that general language competence could be learned once and reused.

GPT-2 scaled the recipe by an order of magnitude, to 1.5 billion parameters (OpenAI, 2019; Radford et al., 2019). The result was qualitatively new: coherent long-form text generation and non-trivial zero-shot

performance. The model could attempt tasks it had never been explicitly trained on, from translation to summarization, purely as a byproduct of scale. OpenAI's decision to stage the model's release over misuse concerns was itself a signal of how far capability had moved.

GPT-3 extended the scaling logic to 175 billion parameters and introduced few-shot learning: a task could be specified through a handful of examples in the prompt, with no retraining at all (Brown et al., 2020). Crucially for our setting, GPT-3 was also the generation that reached the market. Its API opened in beta in June 2020 and to the public in November 2021 (OpenAI, 2020), placing state-of-the-art capability in the hands of any technically capable team. Codex, a GPT-3 derivative fine-tuned on code and deployed as the backbone of GitHub Copilot, brought the models into developer workflows (OpenAI, 2021), and InstructGPT introduced reinforcement learning from human feedback, fine-tuning GPT-3 into models that reliably follow instructions (Ouyang et al., 2022), the immediate technical substrate of ChatGPT.

ChatGPT itself launched on November 30, 2022. Released as a free research preview built on a GPT-3.5-class model, it introduced a key innovation: the intuitive conversational interface (OpenAI, 2022). While OpenAI had circulated powerful language models through academic publications, developer APIs, and niche tooling since 2020, ChatGPT's particular combination of model capability and accessible interface made the technology legible to a vastly wider audience. Many founders, investors, educators, and policymakers only grasped the practical implications of modern LLMs when they could interact with one directly in a browser.

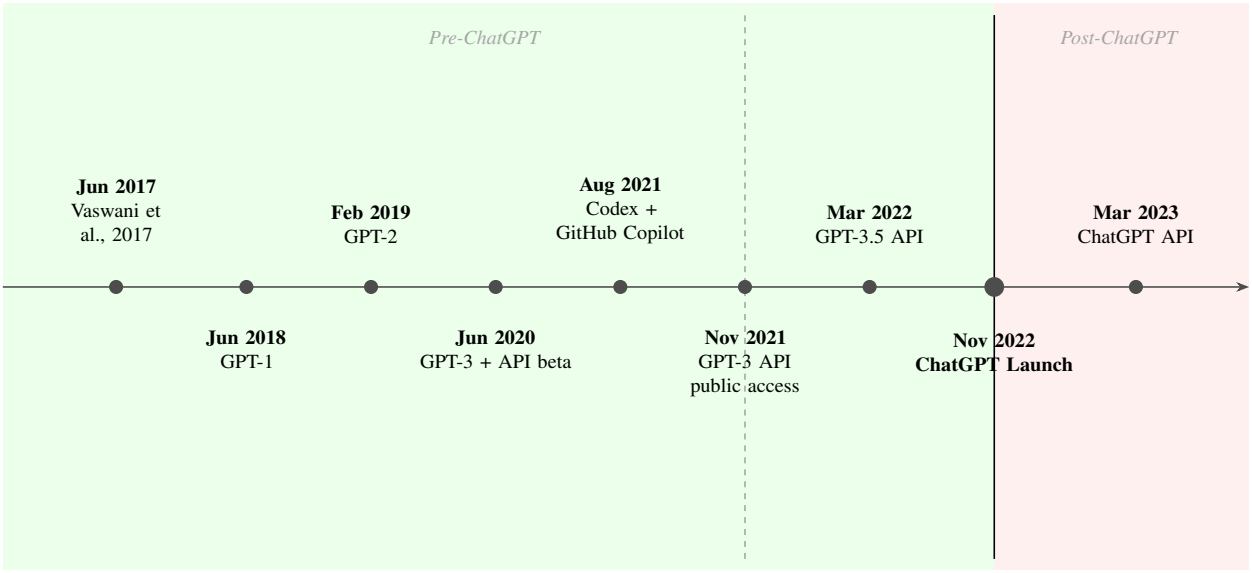
ChatGPT's adoption was immediate: within five days of launch, it had surpassed one million users (Altman, 2022). By January 2023, it had reached an estimated 100 million monthly active users, making it the fastest-growing consumer application in recorded history (UBS, 2023). By October 2025, the platform had grown to more than 800 million weekly active users (Reynolds, 2025).

Figure 1 summarizes the key milestones in the development of the GPT model family leading to the ChatGPT launch.

3.2 ChatGPT as an Economic Shock

The launch of ChatGPT satisfies the conditions for a credible economic shock suitable for causal inference (Eisfeldt & Schubert, 2025). ChatGPT appeared as a sharp, discrete event. Even OpenAI leadership subsequently described the scale of mainstream uptake as surprising. The decision to release a free consumer-facing interface, rather than continuing to serve developers through a paid API, was a strategic pivot whose consequences were not broadly anticipated (including, by many accounts, within OpenAI itself).

Figure 1: The Road to ChatGPT



Notes: Timeline of key developments in the GPT model family leading to the launch of ChatGPT on November 30, 2022. The dashed vertical line marks the public release of the GPT-3 API (November 2021). The solid vertical line marks the ChatGPT launch.

Before November 2022, awareness of generative AI and its underlying capabilities was largely confined to a narrow technical community. The pre-launch user base was, by design, small and specialized. After the launch, ChatGPT diffused at exceptional speed. The trajectory from one million users in five days to 100 million in two months is without precedent in the history of consumer applications, and it ensured that ChatGPT became salient to virtually every sector of the economy in a compressed timeframe.

Consistent with this perspective, a growing empirical literature uses the sudden release of ChatGPT as an exogenous economic shock to study its causal effects on entrepreneurship and labor market behavior.[1]

A useful feature of the setting sharpens what kind of shock the launch was. Successive vintages of the underlying model capability reached the market through the API well before ChatGPT: GPT-3 became publicly accessible in November 2021, and the GPT-3.5 models on which ChatGPT itself was built followed in March 2022 (Figure 1). That access was available precisely to the DL/ML-trained teams we classify as treated, for whom an API posed little technical barrier. We mark the first of these dates with a dashed vertical

[1] Bao et al. (2025) exploit industry variation in exposure to generative AI and show that the launch of ChatGPT increased incorporated entrepreneurial entry among highly exposed STEM workers. Marchesi and Tang (2025) use a difference-in-differences framework based on industry AI compatibility and document a significant post-ChatGPT rise in firm formation, particularly in industries where AI reduces experimentation costs. Yiu et al. (2026) analyze an online labor market and show that freelancers strategically reposition their labor supply following the release of ChatGPT, consistent with a technology-driven shock to labor demand and supply. Zhu et al. (2026) examine the effect on marketing employment and find that marketing headcount declines relative to non-marketing headcount after ChatGPT's release.

line in every event-study figure below; both dates fall within the window of flat differential pre-trends, and no outcome in the paper diverges at either. ChatGPT's dialogue-tuned model was somewhat more capable than its API predecessors, so the comparison is not exact. The launch also bundled an accessible interface with instruction tuning, free consumer availability, and exceptional public salience. The API dates thus function as timing checks rather than clean placebos: they date the onset of the response to the diffusion regime the launch initiated, without isolating the interface's contribution from the bundle.

3.3 Differential Exposure to ChatGPT

Although ChatGPT, in principle, affects all ventures, not all ventures were equally positioned to act on it (Eisfeldt & Schubert, 2025). Firms and workers differ systematically in how much they stood to gain from ChatGPT's arrival, and these differences are measurable ex ante. To implement a difference-in-differences-like design, we must identify a subset of ventures that were more intensely exposed.

Absorptive capacity is "the ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends" (Cohen & Levinthal, 1990). This capacity is cumulative and domain-specific: prior knowledge in a given technical area lowers the cost of acquiring related new knowledge, because learning is associative and self-reinforcing. New information is encoded by establishing linkages with pre-existing concepts, so that "the more objects, patterns and concepts that are stored in memory, the more readily is new information about these constructs acquired." Conversely, absence of such prior knowledge can produce "lockout," a state in which a firm cannot readily access or evaluate external advances even when they become publicly available. ChatGPT's launch made a powerful new technology freely accessible, but making use of it required all three components of absorptive capacity: the ability to recognize what the technology could do, to assimilate its capabilities into existing workflows, and to apply it to concrete business problems.

We argue that deep learning (DL) and machine learning (ML) expertise is a natural proxy for a venture's absorptive capacity with respect to GenAI. Several considerations support this claim. First, DL/ML appears to be a well-defined, path-dependent domain (Gofman & Jin, 2024), suggesting that DL/ML knowledge is not readily substituted by adjacent technical backgrounds; this is the domain-specificity that Cohen and Levinthal (1990) identify as a defining feature of absorptive capacity. In addition, because absorptive capacity is cumulative and self-reinforcing, DL/ML-trained founders had been building relevant knowledge progressively (through prior training, research exposure, and hands-on experience with tools such as GitHub

Copilot) long before ChatGPT's launch. This accumulated foundation enabled them to assess more accurately the commercial potential of the new technology as soon as it became broadly accessible. Founders without DL/ML expertise, by contrast, were likely encountering LLMs for the first time via ChatGPT itself, requiring simultaneous discovery of both the technology and its potential applications: a far steeper learning curve, and the condition under which Cohen and Levinthal (1990) warn that lockout occurs.

These differences map directly onto the three dimensions of absorptive capacity, as summarized in Table 2. Ventures led by DL/ML-trained teams were already tracking LLM progress, could readily interpret the new capabilities, and had existing technical workflows into which GenAI tools could be integrated. Cohen and Levinthal (1990) emphasize that organizational absorptive capacity is not merely the sum of individual capabilities: it depends on whether the team as a whole possesses sufficient shared knowledge to recognize, communicate, and act on new external information. DL/ML expertise in the TMT is our proxy for this organizational foundation: team members with such expertise are positioned to serve as the boundary-spanning nodes through which GenAI knowledge enters and diffuses within the venture, though we observe the expertise, not the diffusion. Ventures without such backgrounds, by contrast, plausibly encountered ChatGPT as a novel object, faced a steeper assimilation curve, and lacked established pathways for integration, the organizational analogue of the lockout condition Cohen and Levinthal (1990) describe.

In our analysis, we distinguish between "AI-ready" ventures, those where key people had DL/ML backgrounds prior to ChatGPT's launch, and "AI-naive" ventures, those with no such background. We assume that AI-ready ventures are more likely to have the ability to recognize, assimilate, and apply GenAI than AI-naive ventures, and thus are more exposed to the shock. We probe the relevance of this assumption in Section 5.

Table 2: Absorptive Capacity by Treatment Status

	AI-Ready *(DL/ML Teams)*	**AI-Naive** *(Other Teams)*
Recognition	Prior knowledge enabled accurate assessment of commercial potential	Lacked the knowledge base to immediately evaluate capabilities
Assimilation	Cumulative DL/ML expertise lowered the cost of integrating new capabilities	Faced the steep learning curve of discovering the technology and its applications
Application	Existing workflows served as boundary-spanning nodes for GenAI diffusion	No established integration pathways

Notes: Classification of treated and control ventures along the three dimensions of absorptive capacity defined by Cohen and Levinthal (1990). AI-Ready ventures have DL/ML expertise prior to ChatGPT's launch; AI-Naive ventures have no such prior expertise.

4 Data

We combine Crunchbase venture data and Revelio Labs employee resume data to construct a firm-month panel of nearly 27,000 U.S. software ventures founded between 2018 and 2021. We supplement this core dataset with website-technology data from BuiltWith, trademark filings from the USPTO, and funding histories and firm-level news coverage from Crunchbase. We describe the sample, the treatment indicator, the construction of each family of outcome variables, and report summary statistics.

4.1 Sample

We construct our dataset by merging venture data from Crunchbase with employee resume data from Revelio Labs. Crunchbase is generally considered the largest publicly available database of U.S. venture activity (Gofman & Jin, 2024), providing information on founding dates, product categories, and funding histories. Revelio Labs collects and structures online profiles, covering over 80% of the U.S. workforce. For each position, the data include job titles, job descriptions, start and end months, and individual-level information on educational background and self-reported skills.

We restrict attention to software ventures founded in the United States between January 2018 and December 2021, to ensure a sufficient pre-shock window between venture birth and the release of ChatGPT. The restriction to software ventures is implemented through an LLM-based procedure, which also classifies ventures into 27 markets.

Applying these restrictions to the Crunchbase universe and requiring the presence of a LinkedIn URL yields 41,475 candidate ventures. We match these to Revelio Labs on the normalized LinkedIn URL and obtain a match rate of approximately 69%. We then drop ventures whose founding years in the two databases disagree by more than two years, arriving at a matched firm list of 27,498 ventures.

For each venture, we construct a firm-month panel beginning in the month of founding and ending in December 2025 or at the venture's closing date as recorded in Crunchbase, whichever occurs earlier. Importantly, the panel frame is constructed entirely from Crunchbase founding and closing information, so that a venture's presence in the panel does not depend on the employment outcomes we study.[2] As a final step, we remove ventures above the 99th percentile of founding-team size and drop ventures observed for fewer than two months. This yields 2,064,493 firm-month observations across 26,873 ventures.

[2]A frame based on observed employment activity yields similar results and is retained as robustness.

4.2 Treatment Indicator

Our primary independent variable is a time-invariant binary indicator for whether a venture was well-positioned to adopt GenAI tools at the time of ChatGPT's launch. Constructing it requires three steps: identifying each venture's top management team (TMT), assessing whether each TMT member possesses DL or ML expertise, and combining the two to classify ventures as *AI-ready*. First, we identify TMT membership from job titles in the Revelio Labs data, classifying as TMT roles founders, C-suite executives, vice presidents, directors, and heads of function.

Second, we assess DL/ML expertise from the self-reported skills in each individual's profile: DL expertise captures neural network architectures, transformer-based models, generative AI, and deep learning frameworks (e.g., PyTorch, TensorFlow); ML expertise captures classical methods such as tree-based models, support vector machines, and traditional NLP tools.

Revelio also reports individuals' educational histories, including degree fields; we classify degree fields into STEM and non-STEM categories, which supports the robustness analysis of Section 6.1. Field information is available for 65.2% of TMT members.

Third, a venture is classified as *AI-ready* if at least one TMT member with DL or ML expertise held both the position and the qualifying leadership title before ChatGPT's launch in November 2022, so that post-shock promotions do not generate treatment. The leadership margin of the classification is verified ex ante and the indicator is time-invariant, mitigating concerns about post-shock endogeneity or strategic sorting; the skills margin is addressed next. Its theoretical basis, the absorptive capacity framework of Cohen and Levinthal (1990), was developed in Section 3. Approximately 23.7% of panel firm-month observations satisfy this baseline criterion; below we refine it into a look-ahead-robust headline definition.

Because resume data are observed after the shock, self-reported skills could embed look-ahead bias: profiles accrete skills over time, and some AI-related vocabulary only entered common usage after ChatGPT's launch. Comparing two vintages of the Revelio data five months apart, 10.4% of individuals flagged with AI skills acquired the flag between vintages: the concern is real though bounded. We therefore exclude the unadjusted skills measure from the treatment set and construct a 2×2 grid of definitions. On the expertise margin, a *vocabulary-cleaned* variant excludes skills whose vocabulary postdates the shock, and a *timing-verified* variant additionally requires the skill to be first observed before December 2022. On the

leadership margin, a *broad* variant uses the full TMT and a *strict* variant restricts to founders and C-suite.[3] The four definitions classify between 17.8% and 4.5% of panel observations as treated. The 17.8% headline share reflects the breadth of the TMT margin in a software-only sample (one qualifying member anywhere among founders, C-suite, vice presidents, directors, or heads of function), while the stricter definitions classify 4.5–12.1%. Our headline treatment is the broad–vocabulary-cleaned definition, which removes the mechanical look-ahead contamination while retaining the full TMT construct our theory motivates; we report the full grid as robustness.

4.3 Workforce

Our primary dependent variables measure how ventures allocate human capital across activities each month. We count an employee as active in a given month if their reported position overlaps with that month, clipping positions to the venture's founding-closing window, and aggregate to the firm-month level. Total active positions is the count of all positions overlapping the month; a small number of individuals hold multiple concurrent positions at the same firm, and a distinct-person denominator yields similar results. Ratios are undefined, not zero, in months without recorded headcount, so share regressions condition on at least one active position.

We classify positions into activities by applying dictionary-based classification to job titles and descriptions, adapting the keyword approach of Wang and Wu (2026). Before classification, we blank job descriptions that repeat verbatim across many positions within the same firm, so that boilerplate company blurbs do not contaminate individual-level classification; matching uses regular expressions with word boundaries and bounded word-proximity, so that co-occurrences across distant sentences do not generate false positives.

We distinguish three types of experimentation, corresponding to distinct modes of experimental learning. *Customer Input* (Exp 1) captures direct elicitation of customer and user needs: interviews, focus groups, surveys, user research, and customer discovery. *Prototyping* (Exp 2) captures inductive, artifact-based market learning associated with lean venture methodology (Ries, 2011): MVP development, prototyping,

[3] A caveat applies to the timing-verified variant: skill dates in Revelio are data-capture dates rather than acquisition dates, and the vendor lost access to skills for most public profiles in May 2022 (Wiggers, 2022), so genuinely pre-shock skills first captured in later crawls appear mechanically as post-shock. The timing-verified definition thus yields a deliberately conservative treated set; our headline definition, which keys on skill content rather than observation timing, is insulated from capture-date gaps, though not from post-launch reporting of pre-existing DL/ML terms. We retain the timing-verified cells precisely because an effect surviving this stacked-against-us definition is the more credible for it.

concept validation, pilot programs, proof-of-concept work, and beta testing. *Controlled Testing* (Exp 3) captures deductive, hypothesis-driven experimentation: A/B and multivariate testing, experimental design, control groups, statistical significance, and feature-flag infrastructure. The activity flags are not mutually exclusive, and a combined experimentation indicator flags any position matching at least one of the three experimentation types. Table 3 reports representative keywords for each category.

Furthermore, we capture efforts associated with scaling the venture, which typically take place as the experimentation process winds down. *Scaling* captures post-product-market-fit growth activities: scaling the business, growth economics (ARR, CAC, LTV), repeatable processes, and market expansion.

Beyond experimentation, we track workforce composition across five mutually exclusive functional categories: Engineering, Design, Product, GTM (sales and marketing), and Admin (operations, finance, legal, and management), with a residual Other group. Each position is assigned to the function with the largest number of matched title and description patterns, with ties broken by a fixed priority order.

For each activity and function, our outcome variable is the *ratio* of related positions to total active positions, which captures relative workforce allocation and measures how GenAI reshapes the composition of the team. Position counts are reported only descriptively, for reasons developed in Section 6.1.

Table 3: Experimentation and Scaling Keywords

Category	Representative Keywords
Customer Input (Exp 1)	Focus Group, Customer/User Interview, Customer Discovery, Customer Development, User Research, Voice of Customer, Concept Test, Customer/User/Market Survey
Prototyping (Exp 2)	MVP, Minimum Viable Product, Prototype, Product/Market Test, Concept or Hypothesis Validation, Pilot Program, Proof of Concept, Beta Test, Soft Launch, Design Sprint, Usability Test, Lean Startup
Controlled Testing (Exp 3)	A/B Test, Split Test, Multivariate Test, Controlled Experiment, Experimental Design, Test/Control Group, Statistical Significance, Hypothesis Test, RCT, P-value, Feature Flag, Canary Release
Scaling	Scale-up, ARR, MRR, LTV, CAC, Unit Economics, Customer Acquisition, Churn, Growth Strategy, Playbook, Standard Operating Procedure, Market Expansion, Localization, Hiring Plan

Notes: Representative keywords used to identify activity-related positions from job titles and descriptions in Revelio Labs data, adapting the approach of Wang and Wu (2026). Classification uses regular-expression matching with word boundaries and bounded proximity on normalized text, after blanking within-firm boilerplate descriptions. Activity flags are not mutually exclusive.

4.4 Technology Stack and Product Development

We complement human-capital measures of experimentation with measures of technological stack and product development. We collect the technological tools each venture deploys on its website from BuiltWith.[4] We measure the overall size of the technology stack and counts of tools in classes aligned with the activity taxonomy: customer-input tools (e.g., Qualtrics, Hotjar), pre-launch tools (waitlist and coming-soon services such as KickoffLabs, the public-facing shell of prototyping activity), controlled-testing tools (A/B-testing and feature-flag infrastructure such as Optimizely and LaunchDarkly), and commerce tools (checkout, payment, and subscription-billing infrastructure such as Stripe, whose adoption marks a product going on sale). We collect the data in November 2025. Because BuiltWith re-crawls sites at a weeks-to-months cadence, adoption in the final months before the November 2025 collection is only partially detected; we end the technology measurement window in October 2025. Because the re-crawl lag can extend to months, coefficients in the final measured months may still partially reflect delayed detection rather than adoption timing. Matching ventures to BuiltWith by website domain covers 24,441 panel ventures; because a queried website with no detected tools is informative while an unqueried one is not, only queried ventures' tool counts are treated as true zeros. These data measure adoption of product and experimentation infrastructure directly, with a signal that does not depend on how positions are described.

Web-technology data come with known standards of use (Stroube & Dushnitsky, 2023, 2025): they are most informative where the website is central to value creation. We therefore define a *website-core* subsample comprising the eleven markets (of twenty-seven) in which the website is the product or storefront, covering roughly a third of the sample, and report all website-technology analyses in Section 6.3 for both samples, placing more interpretive weight on the latter.

In addition, we measure product development using trademark filings, which capture the introduction of new products and brands, the output margin most relevant for early-stage software ventures, for which patenting is rare. We match ventures to USPTO filings and construct a monthly count of applications and an indicator for any filing in the month. Trademark coverage extends through December 2023, providing thirteen post-shock months; we restrict trademark analyses to this window and interpret them as evidence on

[4] BuiltWith data have supported a growing body of entrepreneurship and strategy research. Koning et al. (2022) use BuiltWith's date-stamped detection of A/B-testing tools to measure when startups adopt experimentation and link adoption to growth. Stroube and Dushnitsky (2025) use BuiltWith as a sampling frame, identifying all U.S. domains running Shopify to study whether low-code tools broaden entrepreneurial inclusion. Impink (2025) queries BuiltWith's technology taxonomy to code startups' cloud adoption and measure technological standardization and differentiation. Wright and Saiedi (2025) use BuiltWith data on website language-tool adoption to trace how software startups expand into foreign markets.

the immediate post-shock period rather than the full horizon.

4.5 Financing and Exit

Finally, we measure performance in terms of financing and exit. We measure fundraising activity from the Crunchbase funding-rounds data, classifying every round by type. Dilutive rounds comprise external equity in all its forms (pre-seed through late series, angel, private equity, corporate rounds, equity crowdfunding, and convertible notes); non-dilutive rounds comprise grants, debt, and non-equity assistance; post-IPO instruments and secondary-market transactions are excluded throughout. From these we construct monthly arrival indicators: any external funding (our headline funding outcome), equity funding, and non-dilutive funding.

We complement funding with exit outcomes from the Crunchbase acquisitions and IPO records. We take each venture's first acquisition (excluding leveraged and management buyouts, which are ownership transactions rather than venture exits) and first IPO, dated by announcement, and define exit as the earlier of the two. Because in-window IPOs are rare in a sample of recently founded ventures (twenty events), we analyze the combined exit outcome rather than the two forms separately.

4.6 Summary Statistics

Table 4 reports summary statistics for the main variables used throughout the analysis. The sample is roughly balanced across the pre- and post-ChatGPT periods, and about 18% of panel observations are classified as AI-ready under the headline definition, with the four definitions of the treatment grid ranging from 17.8% to 4.5% and the pre-cleaning diagnostic at 23.7%. Experimentation-related positions account for a modest share of the workforce, concentrated in prototyping-oriented roles. This pattern is consistent with prototyping and customer discovery serving as the primary modes of experimentation in early-stage software ventures, while formal hypothesis-driven testing remains comparatively rare. Engineering and administrative roles account for the largest workforce shares.

Ventures deploy substantial technology stacks, while adoption of dedicated controlled-testing tools is limited. Funding and exit events are infrequent at the monthly level, consistent with ventures raising external capital only a handful of times over their lifecycle.

Table 4: Summary Statistics

	Mean	Median	SD	Min	Max	N
Panel A: Treatment						
Post (ChatGPT Launch)	0.474	0.000	0.499	0.000	1.000	2064493
AI-Ready (broad TMT, raw skills; diagnostic)	0.237	0.000	0.425	0.000	1.000	2064493
AI-Ready (broad TMT, vocab-cleaned; headline)	0.178	0.000	0.382	0.000	1.000	2064493
AI-Ready (broad TMT, timing-verified)	0.064	0.000	0.244	0.000	1.000	2064493
AI-Ready (strict TMT, vocab-cleaned)	0.121	0.000	0.327	0.000	1.000	2064493
AI-Ready (strict TMT, timing-verified)	0.045	0.000	0.208	0.000	1.000	2064493
Panel B: Human Capital						
Experimentation share (combined)	0.028	0.000	0.098	0.000	1.000	1892680
Experimentation share (customer input)	0.009	0.000	0.056	0.000	1.000	1892680
Experimentation share (prototyping)	0.019	0.000	0.082	0.000	1.000	1892680
Experimentation share (controlled testing)	0.003	0.000	0.030	0.000	1.000	1892680
Experimentation count (combined)	0.294	0.000	0.946	0.000	57.000	2064493
Experimentation count (customer input)	0.100	0.000	0.451	0.000	31.000	2064493
Experimentation count (prototyping)	0.182	0.000	0.648	0.000	50.000	2064493
Experimentation count (controlled testing)	0.038	0.000	0.269	0.000	16.000	2064493
Scaling share	0.020	0.000	0.083	0.000	1.000	1892680
Engineering share	0.212	0.125	0.257	0.000	1.000	1892680
Product share	0.018	0.000	0.070	0.000	1.000	1892680
Design share	0.040	0.000	0.117	0.000	1.000	1892680
GTM share	0.187	0.087	0.252	0.000	1.000	1892680
Admin share	0.475	0.438	0.325	0.000	1.000	1892680
Other share	0.068	0.000	0.149	0.000	1.000	1892680
Panel C: Technology and Product Development						
Technology stack (count)	45.542	41.000	36.146	0.000	438.000	1836379
Customer-input tools (count)	0.104	0.000	0.326	0.000	5.000	1836379
Pre-launch tools (count)	0.002	0.000	0.050	0.000	2.000	1836379
A/B testing and feature-flag tools (count)	0.041	0.000	0.212	0.000	6.000	1836379
Commerce tools (count)	0.136	0.000	0.414	0.000	6.000	1836379
Any trademark filed	0.007	0.000	0.081	0.000	1.000	1254466
Trademark filings (count)	0.010	0.000	0.164	0.000	27.000	1254466
Panel D: Funding and Exit						
Funding (any external round)	0.007	0.000	0.086	0.000	1.000	2064493
Funding (equity round)	0.007	0.000	0.081	0.000	1.000	2064493
Funding (non-dilutive round)	0.001	0.000	0.030	0.000	1.000	2064493
Exit (acquisition or IPO)	0.001	0.000	0.023	0.000	1.000	2064373

Notes: Firm-month observations, from founding through closing or December 2025. The AI-Ready indicators follow the 2×2 treatment grid of Section 4, with the broad-TMT, vocabulary-cleaned definition as the headline treatment; the raw-skills row is a pre-cleaning diagnostic, not a treatment definition.

5 Methodology

We illustrate the methodology. First, we provide evidence in support of our design using firm-level news coverage. Then, we present the difference-in-differences design we will be using throughout the analysis.

5.1 Treatment Validation

Before moving to regression analysis, we use firm-level news coverage to establish that AI-ready ventures are differentially associated with GenAI activity after the shock. News-based measures come from Crunchbase press references, which we collect for every venture in the sample, yielding 111,754 articles across the 10,700 ventures with any coverage; the validation exhibits use true article dates, which extend the series through mid-2026. We construct a monthly count of articles and flag articles mentioning AI/ML and generative AI.

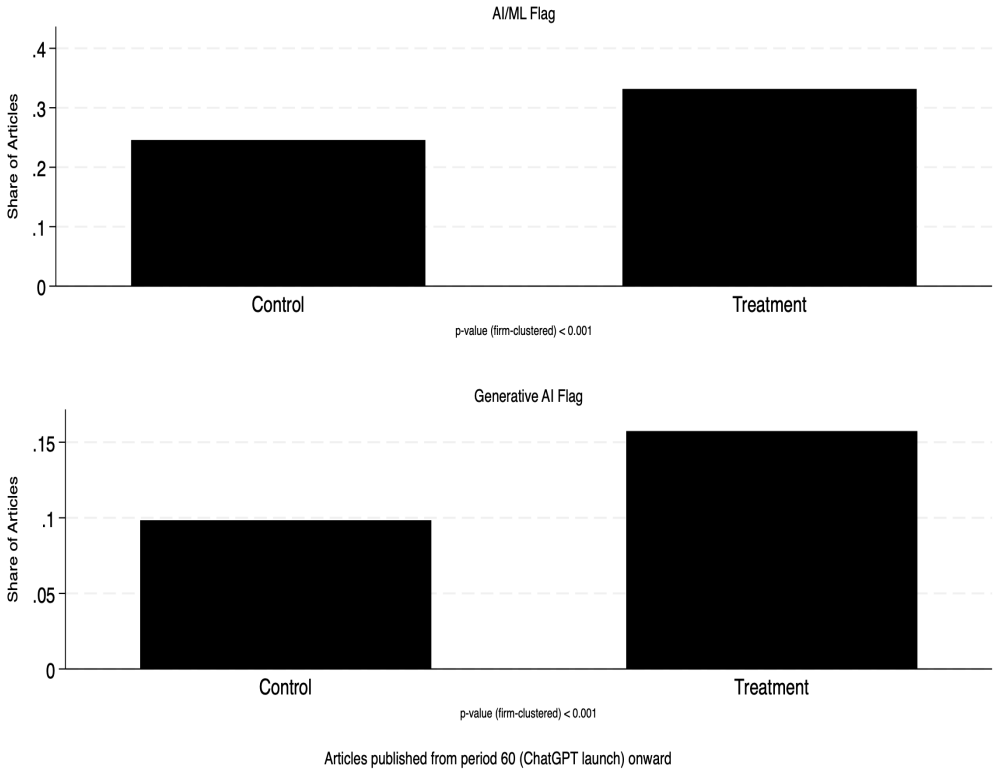
Figure 2 compares treated and control ventures in the post-ChatGPT period: AI-ready ventures receive a substantially higher share of AI-related coverage on both the AI/ML and the generative-AI measure, with differences statistically significant at conventional levels. This comparison is of post-period levels: it establishes that treated ventures are differentially associated with GenAI in post-launch coverage, not that this association reflects a causal response to the launch. The comparison is also conditional on receiving any coverage and, being article-level, weights ventures by article volume; it speaks to the treatment measure's relevance, not to adoption itself.

5.2 Empirical Strategy

Our empirical strategy exploits the unexpected launch of ChatGPT as a common technological shock, combined with cross-sectional heterogeneity in ventures' ex ante capacity to absorb it. This design has several advantages. First, the timing of ChatGPT's release was largely unanticipated, reducing concerns that firms strategically adjusted their hiring or organizational practices in anticipation of the shock. Second, the AI-readiness measure is based on TMT characteristics that predate the shock and are thus plausibly exogenous to post-ChatGPT changes in workforce composition. Third, firm fixed effects absorb all time-invariant differences across ventures (including founding team quality, industry niche, and baseline experimentation culture), while period fixed effects flexibly control for aggregate time trends, macroeconomic conditions, and seasonality common to all firms.

A note on the estimand: control ventures are not untreated in any absolute sense (GenAI was available to all, and control firms plausibly adopted with a lag), so our estimates capture the relative effect of faster absorption rather than of adoption versus no adoption. Because period fixed effects absorb any response common to treated and control ventures, the design is silent on the aggregate effect of GenAI; what the lagged-adoption logic implies is that our estimates understate treated ventures' own adjustment. A related

Figure 2: News Coverage



Notes: Share of news articles flagged as AI/ML-related (top panel) and generative-AI-related (bottom panel) for control and treated (AI-ready) ventures, restricted to articles published from the ChatGPT launch (November 2022) onward. p-values below each panel are from article-level regressions of the flag on treatment status, with standard errors clustered at the firm level.

point concerns what shock the dynamics identify: the post-launch window folds in the successive capability improvements that followed ChatGPT, which AI-ready teams plausibly also absorbed faster, so the coefficients trace the response to the diffusion regime the launch initiated rather than to the November 2022 interface alone. The API timing checks of Section 3.2 date the onset of that regime; they do not isolate the interface's contribution from the capability gains that followed within it. A note on dynamics: the outcomes are workforce shares, which adjust through hiring and attrition flows, so even an immediate change in hiring policy produces a gradual, cumulating share response rather than a jump on impact, consistent with the roughly eight-month onset delay and the growth of the estimated effects through the post window.

Throughout the main analysis, AIReady_i denotes our headline treatment definition: the broad TMT margin combined with vocabulary-cleaned skills (Section 4). This choice is measurement-motivated. The vocabulary cleaning removes the mechanical look-ahead contamination we documented in the skills data—profile terms that only entered common usage after ChatGPT's launch—while retaining the full TMT construct

that our theoretical framework motivates. Under this definition, 17.8% of panel observations are treated. We report results across all four definitions of the treatment grid as robustness, and they tell a consistent story. We estimate two specifications for each outcome.

The first is a dynamic event study that allows the treatment effect to vary freely across periods:

$$Y_{it} = \sum_{\substack{\tau=-59 \\ \tau\neq -1}}^{36} \beta_\tau \cdot (\text{AIReady}_i \times \mathbb{1}\{t=\tau\}) + \alpha_i + \delta_t + \varepsilon_{it}$$

where $\mathbb{1}\{t=\tau\}$ are period dummies defined relative to the ChatGPT launch, with $t=-1$ (November 2022) as the omitted reference period (ChatGPT launched on November 30, 2022, so November is the last essentially untreated month), and α_i and δ_t are firm and period fixed effects. Pre-treatment coefficients ($\tau<-1$) provide a direct test of the parallel trends assumption; post-treatment coefficients trace the dynamics of the response. Standard errors are clustered at the firm level throughout.

The second is the standard difference-in-differences specification that collapses the post-period into a single indicator:

$$Y_{it} = \beta \cdot (\text{AIReady}_i \times \text{Post}_t) + \gamma' \mathbf{X}_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

where Post_t equals one for all periods from December 2022 onward. The coefficient β is the DiD estimand, capturing the average post-shock change in the outcome for AI-ready ventures relative to the control group. In the static specification, $\mathbf{X}_{it}$ is a vector of contemporaneous market-level controls: the log counts of (i) active control ventures, (ii) control ventures raising an equity round, and (iii) positions at control ventures, in the venture's market. Computing these aggregates over control ventures only ensures that treated ventures' own post-shock behavior cannot enter the controls. The dynamic specification omits the controls, so that the pre-trend evidence reflects the raw differential paths rather than covariate-adjusted ones; static estimates include the controls throughout. The pre-trend evidence thus speaks to the unconditional parallel-trends assumption; we note this mismatch explicitly rather than re-estimating the dynamics with controls. Because the static coefficient averages the post-period against the full pre-period, it folds in the early differential convergence visible in distant event time. The dynamic estimates, anchored to the final pre-launch months, are the timing-clean evidence, and we report the static coefficient as the conventional summary rather than as an independently identified magnitude.

6 Analysis

We examine the impact of ChatGPT on venture behavior and outcomes. We begin with human capital, the core of the analysis: experimentation, the focal outcome, together with its scaling contrast (Section 6.1), followed by functional team composition (Section 6.2), which characterizes the breadth of the workforce response. We then move downstream, examining website-technology adoption (Section 6.3), product introduction measured through trademark filings (Section 6.4), and financing and exit (Section 6.5). Throughout, we present dynamic event studies alongside static difference-in-differences estimates, reading the latter jointly with the former.

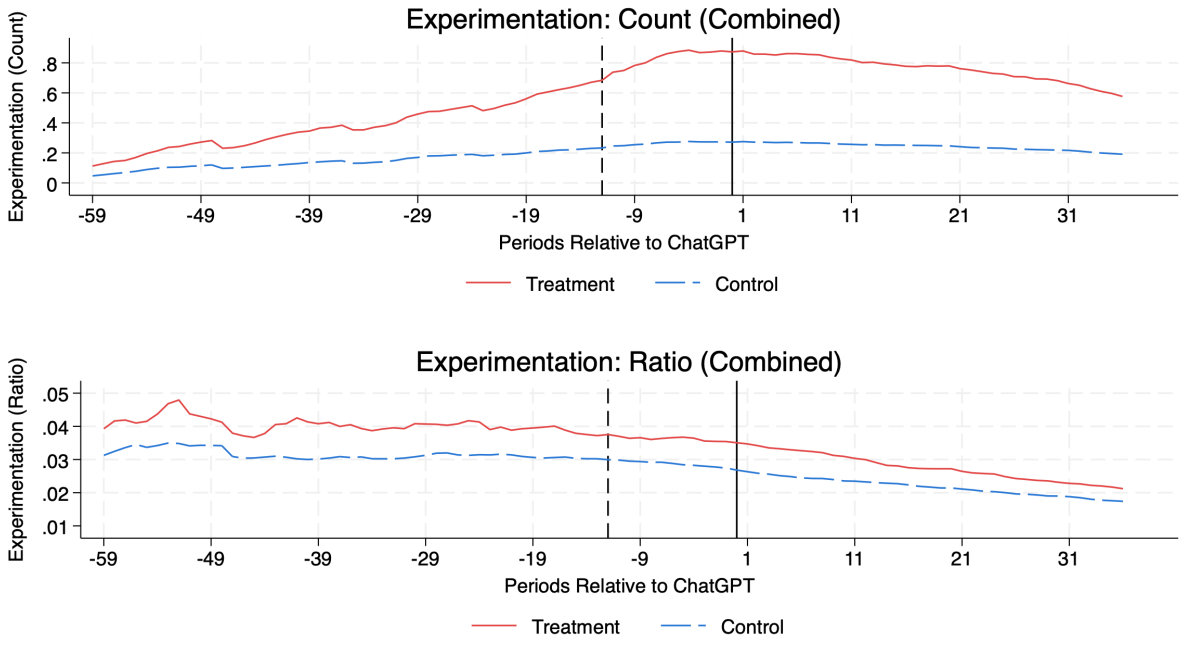
6.1 Experimentation

We begin with the raw data. Figure 3 plots the average count and share of experimentation-related positions over time, separately for treated and control ventures, centered on the ChatGPT launch. AI-ready ventures maintain a higher absolute level of experimentation hiring throughout, reflecting their larger average team size.[5] In share terms, the two groups track each other closely through the pre-period; following the launch, the experimentation share declines more steeply among AI-ready ventures, with the gap opening gradually and widening through the end of the sample. This visual divergence, concentrated in the post-shock period, foreshadows the estimates below. The raw series also bound what the denominator can explain: the treated experimentation count itself peaks near the launch and declines in absolute terms over the post window (Figure 3, top panel), so the share decline is not a pure artifact of accelerating denominators. As noted, the count's diverging pre-trends still disqualify it from carrying causal weight.

Figure 4 presents the dynamic event-study estimates for the experimentation share. Pre-treatment coefficients are noisy in the earliest periods, when few firms are observed at those event times, but settle tightly around zero over the final twenty months before the launch, with no discernible trend. The coefficients remain near zero for roughly the first eight months after the launch and then decline steadily, reaching approximately -0.45 percentage points by the end of the sample window. The raw experimentation share declines secularly in both groups (Figure 3); the event study nets out this common component through the period fixed effects. What remains is a differential pre-decline visible in early event time (treated ventures' share falling faster than controls') that flattens out over the final twenty months before the launch, followed

[5] As the raw data display diverging pre-trends, we do not examine count-based variables in regression analysis.

Figure 3: Experimentation Trends by Treatment Status

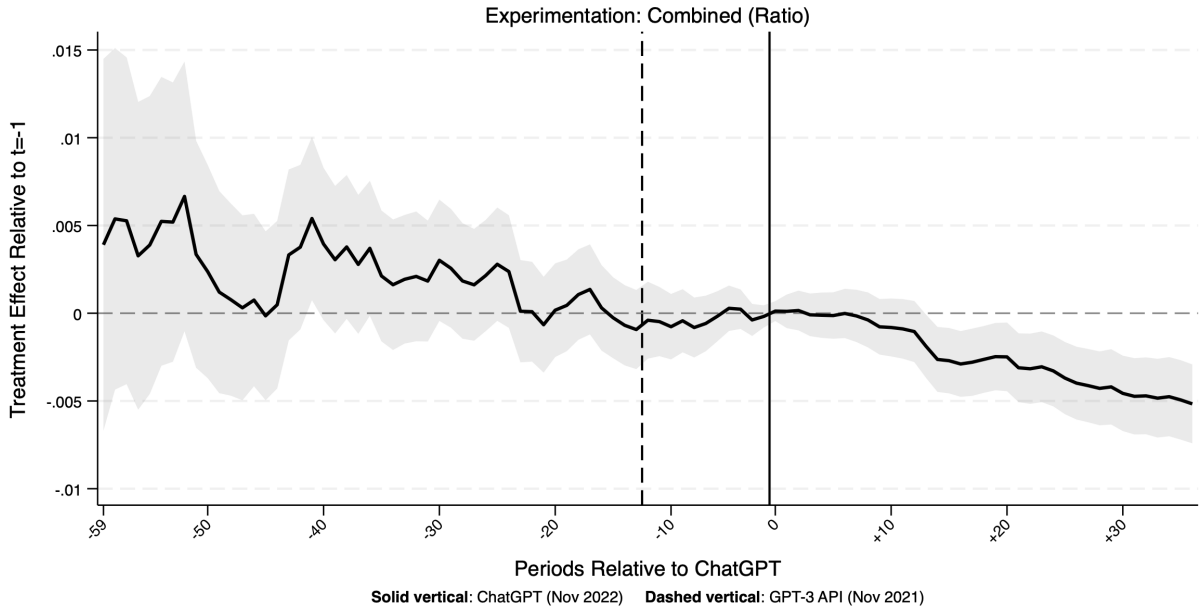


Notes: Average experimentation-related position counts (top panel) and workforce shares (bottom panel) by treatment status, plotted against periods relative to the ChatGPT launch. Solid lines represent AI-ready ventures; dashed lines represent control ventures. The dashed vertical line marks the public release of the GPT-3 API (November 2021); the solid vertical line marks the ChatGPT launch (November 2022).

by a post-launch divergence whose gradual onset is consistent with the stock-adjustment dynamics noted in Section 5.

Figure 5 disaggregates the dynamic estimates by experimentation type, with the three event studies stacked on a common scale so amplitudes are directly comparable. The decomposition sharpens the static result: the aggregate decline is a prototyping phenomenon. For customer input gathering (top panel), coefficients sit tightly on zero throughout the sample window: there is no pre-trend and no post-launch response, only a slight and statistically indistinguishable drift at the very end. Prototyping (middle panel) reproduces the aggregate dynamics almost exactly. Distant pre-period coefficients are elevated relative to the reference month (the same secular differential decline visible in the combined series) but settle around zero over the final year before the launch. The coefficients then hold near zero for roughly the first eight months post-launch before declining steadily to approximately -0.4 percentage points by the end of the window, with confidence bands clear of zero from mid-2024 onward. Controlled testing (bottom panel) is the most

Figure 4: Dynamic DiD Event Study: Experimentation Share

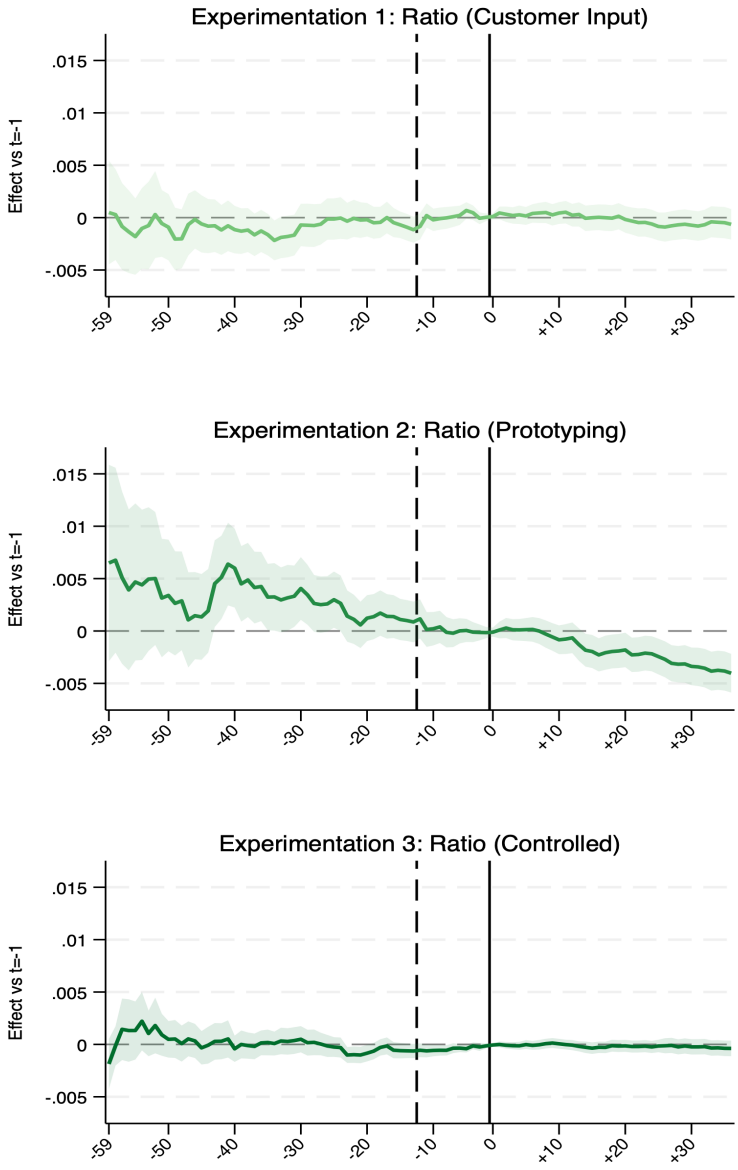


Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the workforce share of experimentation-related positions, with 95% confidence intervals. The omitted reference period is $t = -1$ (November 2022). The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

precisely estimated zero of the three: the share is small and stable, and the post-launch path never departs from the reference level.

Table 5 reports the static DiD estimates for the workforce share. The combined experimentation share declines by 0.35 percentage points ($\hat{\beta} = -0.0035$, $p < 0.01$; Column 1), roughly a 13% decline relative to the 2.8% baseline share, or equivalently about one-third of an experimentation position per hundred workforce positions. The decomposition across experimentation types is informative: the effect is concentrated in prototyping-oriented experimentation (-0.004, $p < 0.01$; Column 3), while customer input gathering (Column 2) and controlled testing (Column 4) show precise zeros. The compositional shift away from experimentation, in other words, is concentrated in inductive, artifact-based market learning (MVP building, piloting, concept validation) rather than in customer research or formal hypothesis testing. Because the activity flags are not mutually exclusive, Columns 2–4 do not additively decompose Column 1; the statement is about where the detectable marginal effect sits.

Figure 5: Dynamic DiD Event Study by Experimentation Type



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification, estimated separately for the workforce share of each experimentation type: customer input gathering (top), prototyping (middle), and controlled testing (bottom), with 95% confidence intervals. Panels share a common vertical scale. The omitted reference period is $t = -1$ (November 2022). The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Table 5: Difference-in-Differences: Experimentation Workforce Share

	(1) Combined	(2) Customer Input	(3) Prototyping	(4) Controlled
ChatGPT × AI-Ready	-0.004***	0.000	-0.004***	0.000
	(0.001)	(0.001)	(0.001)	(0.000)
Observations	1892671	1892671	1892671	1892671
Firm FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R-squared	0.00035	0.00026	0.00031	0.00004

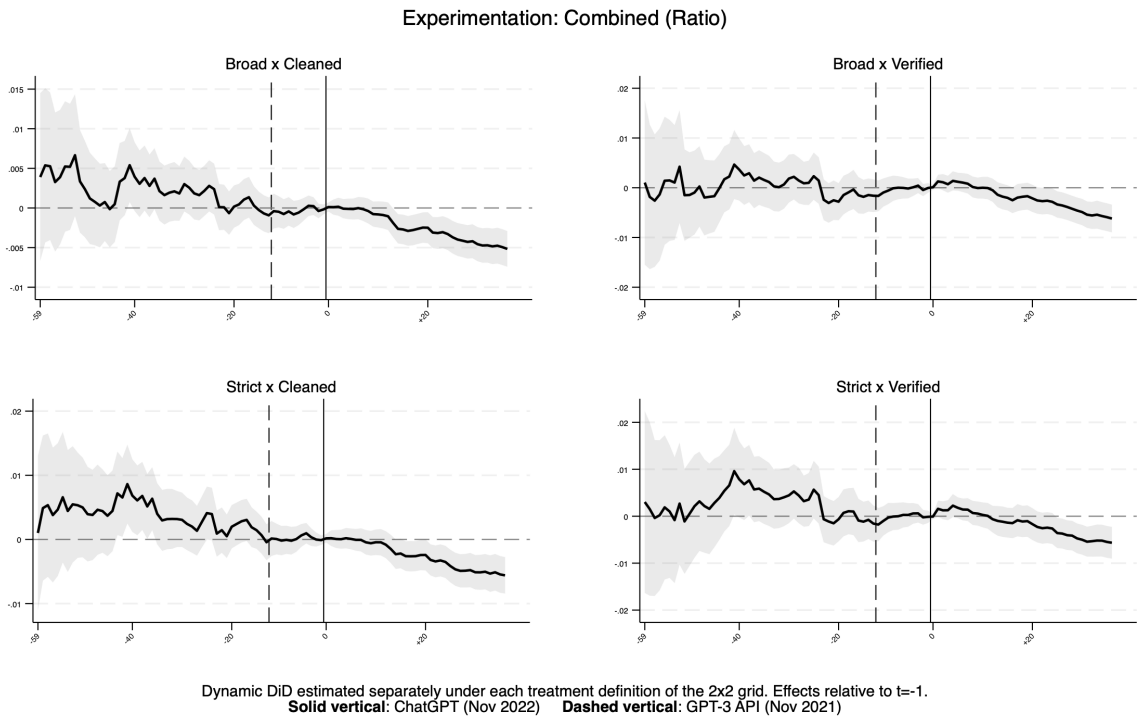
Notes: Difference-in-differences estimates. Dependent variables are workforce shares: the ratio of experimentation-related positions to total active positions, for combined experimentation (Column 1), customer input gathering (Column 2), prototyping (Column 3), and controlled testing (Column 4). The estimation sample excludes firm-months with zero measured positions, for which a workforce share is undefined (171,813 of the 2,064,493 panel firm-months); no ventures are excluded, and the estimator additionally drops a handful of singleton observations. All columns include contemporaneous market-level controls: the log counts of active control ventures, of control ventures raising an equity round, and of positions at control ventures in the venture's market. AI-Ready equals 1 for ventures with at least one TMT member holding vocabulary-cleaned DL/ML expertise prior to ChatGPT's launch. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

To check the robustness to our treatment construction, Figure 6 repeats the dynamic event study under each of the four treatment definitions, crossing broad and strict TMT margins with vocabulary-cleaned and timing-verified skills. The pattern is uniform: recent pre-trends are flat and the post-launch decline appears in every cell, with point estimates between -0.002 and -0.005 and statistical significance in three of four cells. The timing-verified definitions, which shrink the treated group to as little as 4.5% of panel observations, remain the least precise: the strict cell is significant only at the ten-percent level, and the broad cell is same-signed but insignificant. The result is not an artifact of any particular treatment definition.

Table 6 reports the corresponding static estimates. Column 1 repeats the headline definition; Columns 2–4 report the alternative cells of the grid. The point estimate is negative in every column. Column 3, the strict vocabulary-cleaned definition, is significant at the 1% level; Column 4, the strict timing-verified definition, retains significance only at the 10% level; and Column 2, the broad timing-verified definition, is same-signed but insignificant, consistent with the precision loss under the timing-verified definitions documented above.

Beyond these design-based checks, Table 7 subjects the headline estimate to four further tests. Column 1 addresses the concern that the treatment conflates ventures positioned to *use* generative AI with ventures *building* AI products, for which ChatGPT was also a demand and competitive shock. Excluding every venture whose Crunchbase categories indicate an AI/ML product market (a deliberately over-inclusive filter that removes 15.8% of the sample, including a third of treated ventures) leaves the estimate, if anything,

Figure 6: Dynamic DiD Event Study by Treatment Definition: Experimentation Share



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the combined experimentation share, estimated separately under each treatment definition of the 2×2 grid: {broad, strict TMT} × {vocabulary-cleaned, timing-verified skills}. The top row varies the skills margin under the broad TMT definition; the bottom row under the strict definition. Effects relative to $t = -1$, with 95% confidence intervals. All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Table 6: Grid Robustness: Experimentation Share Across Treatment Definitions

	(1) Broad-Cleaned	(2) Broad-Verified	(3) Strict-Cleaned	(4) Strict-Verified
ChatGPT × AI-Ready	-0.004***	-0.002	-0.005***	-0.003*
	(0.001)	(0.001)	(0.001)	(0.002)
Observations	1892671	1892671	1892671	1892671
Firm FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
R-squared	0.00035	0.00023	0.00039	0.00025

Notes: Difference-in-differences estimates of the combined experimentation workforce share under each of the four treatment definitions: broad vs. strict TMT (founders and C-suite only), crossed with vocabulary-cleaned and timing-verified DL/ML skills. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

slightly larger (-0.004, $p < 0.01$). Column 2 tightens the control group to ventures with at least one STEM-degreed TMT member pre-launch, so treated ventures are compared against technical, but not DL/ML-trained, founding teams. The estimate falls to -0.002 and retains significance only at the ten-percent level, indicating that part of the headline effect reflects a general technical-team difference rather than DL/ML expertise specifically; the confidence interval comfortably includes the baseline, and the non-STEM classification is conservative (see notes). The pattern across control groups is itself informative: within the education-covered sample, the estimate is -0.0048 against non-STEM control teams, -0.0028 in the pooled covered sample, and -0.0019 against STEM-only controls, a monotone gradient in the control group's expected absorptive capacity. This gradient is exactly the pattern implied if STEM-trained but non-DL/ML teams were themselves partially treated, absorbing GenAI faster than non-technical teams though more slowly than DL/ML-trained ones, which attenuates the treated–STEM contrast. We therefore read Column 2 as a more demanding comparison, consistent with partial treatment of STEM control teams, rather than as evidence against the DL/ML mechanism. The gradient is also compatible with composition differences or group-specific trends, so we do not claim it as a demonstrated bound. Column 3 redefines treatment to require deep-learning expertise specifically (the skill set most directly implicated in generative-model absorption), with classical-ML-only ventures remaining in the control group; the estimate is indistinguishable from the baseline. Column 4 adds founding-cohort-by-period fixed effects, comparing ventures founded in the same year within the same calendar month and thereby purging cohort-specific lifecycle trends in experimentation; the estimate is unchanged. The compositional shift away from experimentation is thus not driven by AI-product ventures, lifecycle composition, or any particular skill margin of the treatment; the one qualification is that against exclusively technical control teams the estimate is roughly half-sized and only marginally significant.

Further, we ask if the decline in the experimentation share reflects a broader reorientation away from search activities, for instance a wholesale pivot from exploration to growth. Scaling activities provide the designed contrast. Positions devoted to scaling (growing the customer base, building repeatable processes, managing growth economics) represent the execution-oriented counterpart to experimentation in the venture development cycle. If AI-ready ventures were simply maturing faster into growth mode, their scaling share should rise as their experimentation share falls.

Figure 7 presents the dynamic estimates. The scaling share shows no movement for roughly the first half-year after the launch, followed by a mild drift reaching about -0.3 percentage points by the end of the window, later in onset and smaller in magnitude than the experimentation decline. In static terms, the scaling

Table 7: Additional Robustness: Experimentation Workforce Share

	(1) Excl. AI-Product	(2) STEM TMT	(3) DL-Only Treatment	(4) Cohort × Period FE
ChatGPT × AI-Ready	-0.0040***	-0.0019*	-0.0040***	-0.0036***
	(0.0013)	(0.0011)	(0.0011)	(0.0010)
Observations	1593090	1164391	1892671	1892671
Firm FE	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Cohort × Period FE	No	No	No	Yes
Sample	Excl. AI-product	STEM TMT (covered)	Full	Full
Treated ventures	3,138	4,283	3,692	4,779
R-squared	0.0003	0.0004	0.0004	0.0003

Notes: Difference-in-differences estimates of the combined experimentation workforce share; the baseline estimate appears in Table 5, Column 1. Column 1 excludes ventures whose Crunchbase category groups or tags match AI/ML product markets (artificial intelligence, machine learning, natural language processing, computer vision, generative AI, and related terms; the filter is deliberately over-inclusive). Column 2 restricts the sample to ventures with at least one TMT member holding a STEM degree, with leadership title and joining date predating the launch; restricting only on education-field coverage (without the STEM requirement) yields -0.0028 (s.e. 0.0011), while comparing treated ventures against covered non-STEM control teams yields -0.0048 (s.e. 0.0013). Column 3 requires deep-learning expertise specifically for treatment; ventures with ML-only TMT expertise remain in the control group. Column 4 adds founding-year-by-period fixed effects. All columns include contemporaneous market-level controls (see Table 5 notes). All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

share is unchanged: the DiD coefficient is -0.001 (s.e. 0.001) and statistically indistinguishable from zero, with controls included, a null on a baseline share (2.0%) comparable to experimentation's.

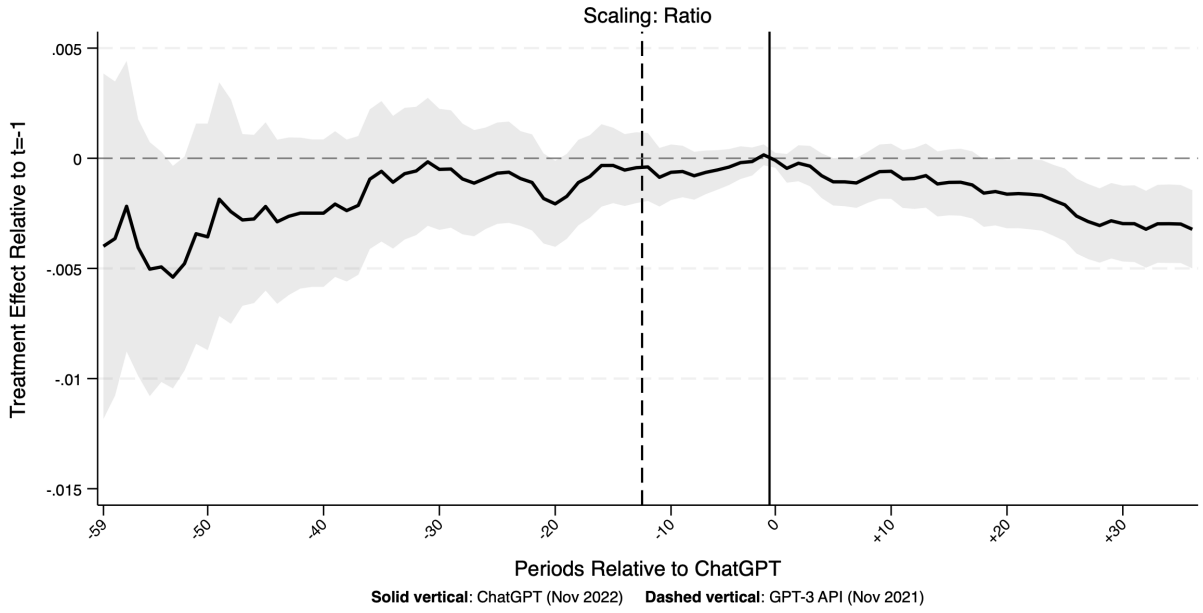
Together with the experimentation results above, this pattern weighs against the simple lifecycle story. We find no indication that AI-ready ventures swapped exploration for execution across the board: the compositional shift is concentrated in experimentation (prototyping in particular), while the scaling share holds steady near the launch and softens only slowly and much later. Whatever GenAI changed for these firms, it changed the role of experimentation labor first and most.

6.2 Functions

We next ask where in the organization the compositional shift occurs, estimating the same battery on the workforce share of each mutually exclusive functional category: Engineering, Product, Design, GTM (sales and marketing), Admin, and Other. Shares sum to one by construction, so the estimates jointly describe how AI-ready ventures rebalanced their teams. This decomposition also reveals where the offsetting reallocation, if any, occurs.

Figure 8 presents the corresponding dynamic event studies, one panel per function. The headline

Figure 7: Dynamic DiD Event Study: Scaling Share

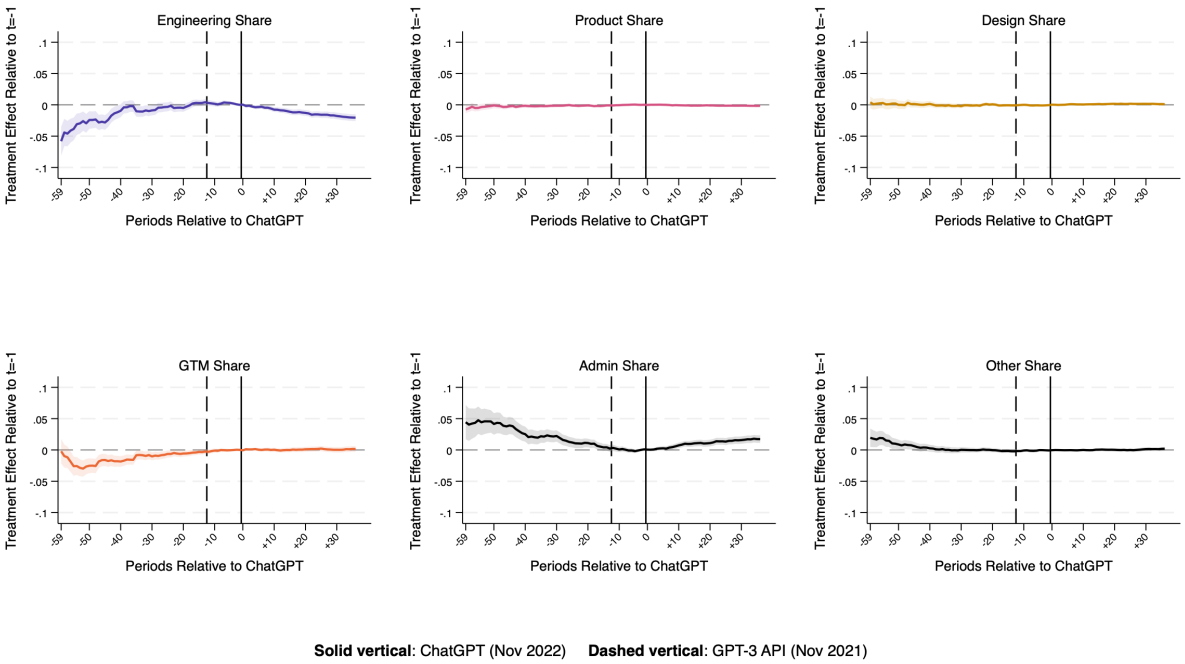


Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the scaling workforce share, with 95% confidence intervals. The omitted reference period is $t = -1$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

finding is a decline in the Engineering share (-0.008, $p < 0.01$), whose dynamics support a causal reading: the Engineering panel shows flat pre-trends over the two years preceding the launch followed by a steady post-launch decline that grows to -0.020 (two percentage points) by the end of the window. The activity flags of Section 4 cut across these mutually exclusive functional categories, and prototyping-flagged positions plausibly sit disproportionately within Engineering. The Engineering decline is therefore best read as the functional-side expression of the reallocation documented in Section 6.1, a related margin of one compositional shift rather than a fully independent second finding. The remaining functions illustrate why we insist on the dynamic evidence before interpreting any static coefficient. The GTM share, for example, carries a positive and significant static estimate, seemingly a shift toward commercialization, but its event-study panel reveals that treated ventures' GTM share was converging upward from below throughout 2019–2021 and is entirely flat after the launch: the static coefficient reflects pre-period convergence, not a post-shock response, and we do not interpret it causally. Product, Design, Admin, and Other shares show no significant

static effects; the Admin panel exhibits its own pre-period drift that likewise precludes causal attribution of its (null) coefficient.

Figure 8: Dynamic DiD Event Studies: Functional Shares



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for each functional workforce share, with 95% confidence intervals. Effects relative to $t = -1$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Table 8 reports the share estimates.

Read together, the three outcome families paint a precise picture of the reallocation. After ChatGPT's launch, AI-ready ventures shifted workforce composition away from the technical core of the search process—prototyping-oriented experimentation and engineering—while their scaling posture held steady. Arithmetically the offset must land somewhere, but no function offers a causally attributable destination: the statically significant GTM rise reflects pre-period convergence already under way and is flat after the launch; the Admin share is the principal late-window accounting counterpart in Figure 8, but its own pronounced pre-period drift precludes causal attribution; and the remaining shares are null. The destination of the reallocation is therefore not identified: we can say where the workforce share fell, not where it went.

Table 8: Difference-in-Differences: Functional Workforce Shares

	(1) Engineering	(2) Product	(3) Design	(4) GTM	(5) Admin	(6) Other
ChatGPT × AI-Ready	-0.008***	0.000	0.001	0.007***	-0.000	-0.000
	(0.002)	(0.001)	(0.001)	(0.002)	(0.003)	(0.001)
Observations	1892671	1892671	1892671	1892671	1892671	1892671
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.00075	0.00004	0.00010	0.00051	0.00087	0.00017

Notes: Difference-in-differences estimates. Dependent variables are functional workforce shares (mutually exclusive categories summing to 1): Engineering (Column 1), Product (Column 2), Design (Column 3), GTM (Column 4), Admin (Column 5), and Other (Column 6). All columns include contemporaneous market-level controls (see Table 5 notes). As discussed in the text, static coefficients are interpreted jointly with the per-function pre-trends evidence in Figure 8. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6.3 Technology Stack

We examine five website-technology outcomes (overall stack size and the four tool classes of Section 4: customer-input, pre-launch, controlled-testing, and commerce tools) on both the full covered sample and the website-core subsample where these data are most valid; we describe both and rest interpretation on the latter. We proceed in three steps: the overall size of the technology stack, the three experimentation tool classes, and commerce infrastructure. Interpretation follows the same discipline as the workforce analysis: static estimates are read jointly with their dynamics.

Table 9 reports the website-core static estimates for the five website-technology outcomes; full-sample counterparts are quoted below. We begin with the overall technology stack. AI-ready ventures expand their stacks after the launch ($+0.077$ log points, $p < 0.05$, in website-core markets; $+0.056$, $p < 0.01$, in the full sample): the descriptive picture is one of broad-based digital buildout. Absent a dynamic event study for overall stack size, we read this level result descriptively rather than causally: without dynamics it may continue a pre-existing differential rather than mark a launch response. The question the remaining outcomes answer is *where* this buildout occurs—and, in particular, whether any of it flows into the infrastructure that supports experimentation.

Next, we zoom in on the three experimentation tool classes, whose dynamic event studies appear in Figure 9. Descriptively, AI-ready ventures carry systematically richer experimentation tooling (the full-sample static estimates for customer-input and controlled-testing tools are positive and precise), but the event studies reveal that these differences predate the shock: both classes converge upward from below

throughout the pre-period, and their post-launch paths are flat to mildly declining. The same convergence generates the positive static coefficients for these classes in Table 9. As with the GTM share in Section 6.2, we therefore read the experimentation-tooling coefficients as descriptive of who AI-ready ventures are, not of what ChatGPT changed. The causal statement is the dynamic one: *adoption of experimentation tooling did not accelerate after the launch*. There is no complementary buildout, paths are flat to mildly declining, and in the website-core sample, where an A/B test is a first-order product decision, the controlled-testing path drifts mildly negative over the post window.

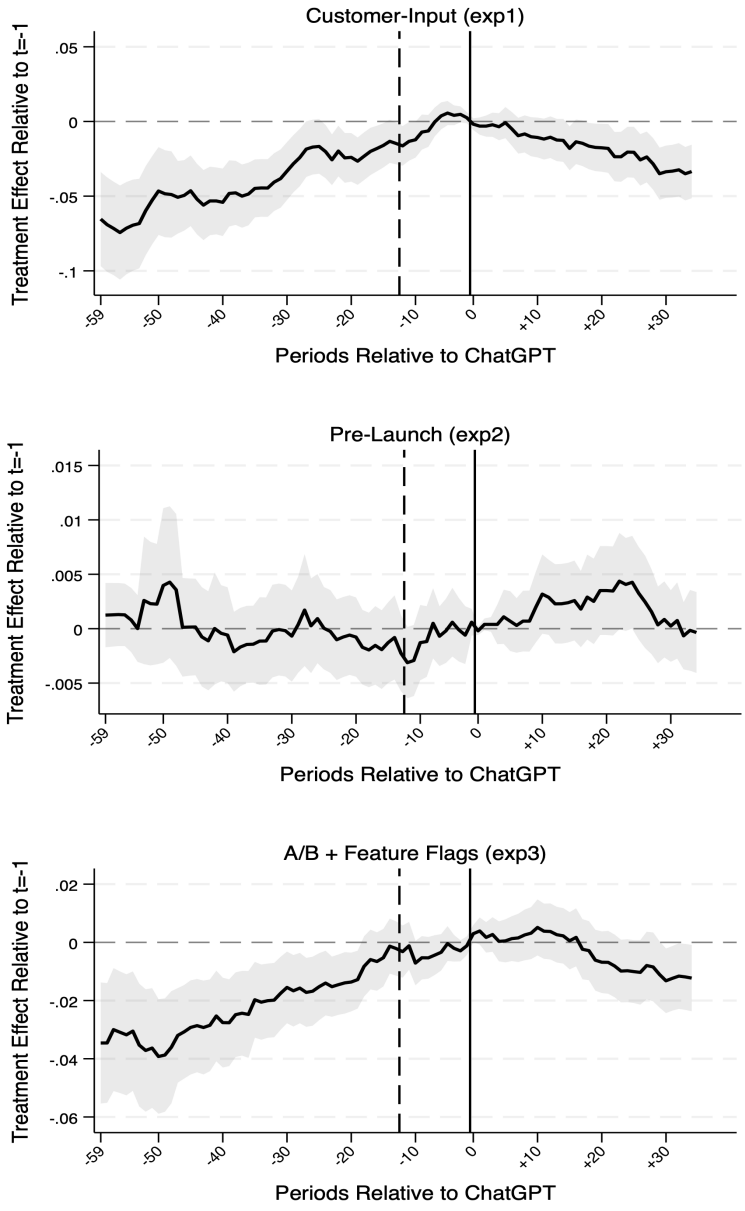
Pre-launch tools, the public shell of prototyping activity, show a precise zero in the full sample; the website-core coefficient is statistically significant but uninterpretable: only 220 ventures ever carry such a tool, the tools are detectable only in their public-facing shell, and the magnitude is negligible. We therefore read both as detection-limited, per the standard of use noted in Section 4. This evidence weighs against the expansionary readings of both the efficiency and precision channels: their shared prediction, complementary investment in experimentation infrastructure, finds no support, now from a data source entirely independent of how positions are described. It does not, however, separate a fixed-quantity efficiency reading (the same experiments, run more cheaply) from substitution (experiments displaced by non-experimental learning); both imply exactly this footprint, a distinction we return to in Section 7.

Finally, we examine commerce infrastructure: the exception, and it points in a specific direction. Adoption of checkout, payment, and subscription-billing tools rises after the launch ($+0.013$, $p < 0.01$, full sample; $+0.011$, same-signed but insignificant, website-core), on flat recent pre-trends, with the full-sample post path climbing steadily through the end of the window (Figure 10). The website-core estimate is imprecise, so we present the result as suggestive rather than definitive. But commerce is the one tool class whose adoption moves *up* while experimentation tooling stands still, and that it is stronger in the full sample than in website-core markets is itself informative: born-digital storefronts carry commerce infrastructure from the outset, so the margin of adjustment sits with ventures digitizing their sales motion. Read together with the workforce results, the technology evidence is suggestive of a reallocation away from building experiments; the commerce margin hints at, but does not identify, a commercialization destination.

6.4 Product Development

Trademark applications provide the formal counterpart to the technology evidence: they record the introduction of new products. Two features of the setting discipline expectations. Coverage ends in December 2023,

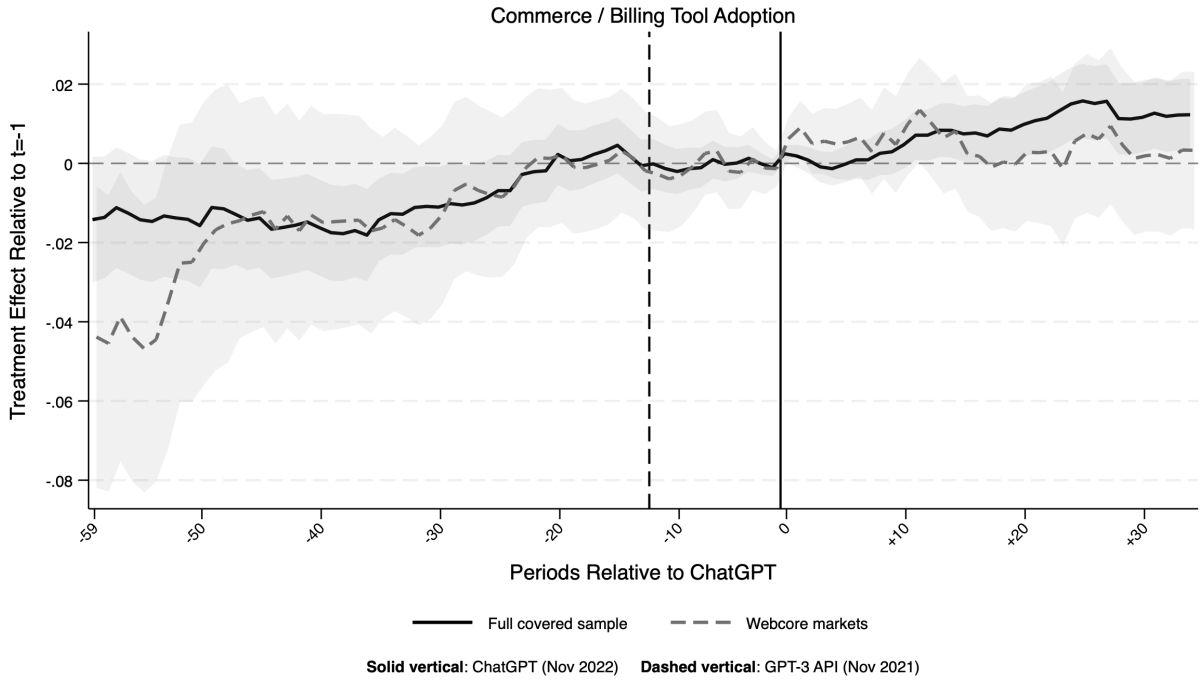
Figure 9: Dynamic DiD Event Studies: Experimentation Tooling (Website-Core Markets)



Sample: website-core markets. **Solid**: ChatGPT (Nov 2022). **Dashed**: GPT-3 API (Nov 2021)

Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the log count of the three experimentation tool classes—customer-input tools, pre-launch tools, and A/B-testing and feature-flag tools—on the website-core subsample, with 95% confidence intervals. Y-axis scales differ across panels. Effects relative to $t = -1$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Figure 10: Dynamic DiD Event Study: Commerce Tool Adoption



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the log count of commerce tools (checkout, payment, and subscription-billing infrastructure), estimated on the full covered sample and the website-core subsample and overlaid, with 95% confidence intervals. Effects relative to $t = -1$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Table 9: Difference-in-Differences: Website Technology Adoption (Website-Core Markets)

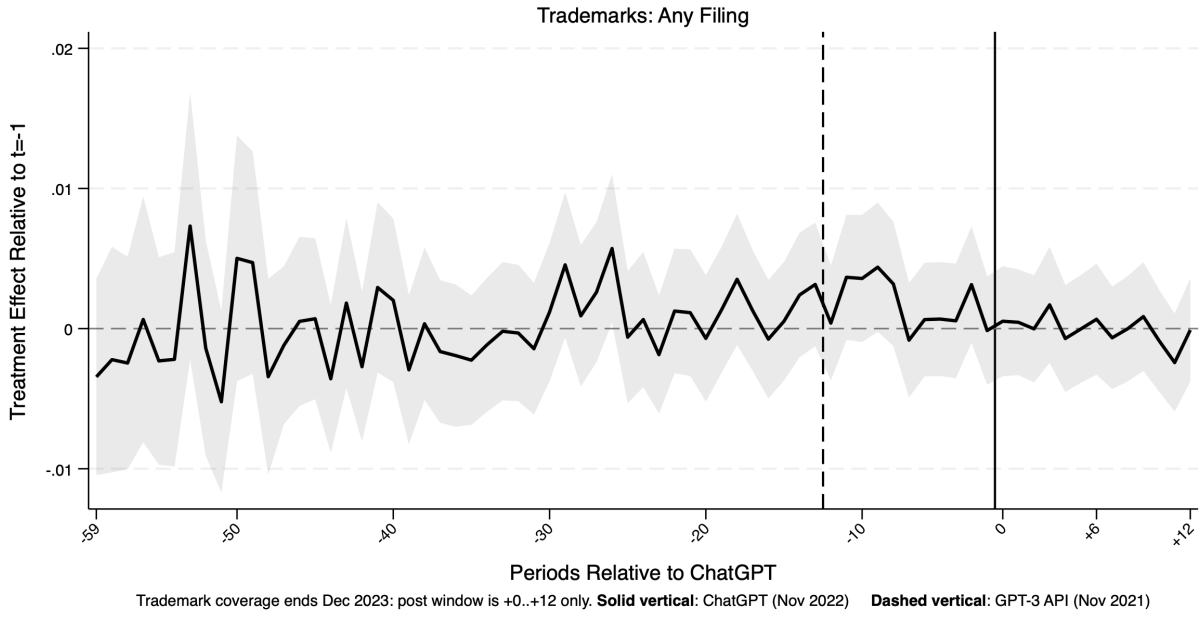
	(1) Tech Stack	(2) Customer-Input Tools	(3) Pre-Launch Tools	(4) A/B + Feature Flags	(5) Commerce Tools
ChatGPT × AI-Ready	0.077**	0.006	0.002**	0.009**	0.011
	(0.033)	(0.006)	(0.001)	(0.004)	(0.007)
Observations	619937	619937	619937	619937	619937
Sample	Restricted	Restricted	Restricted	Restricted	Restricted
Firm FE	Yes	Yes	Yes	Yes	Yes
Period FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Ever-carrier firms	8,243	2,283	220	963	3,060
R-squared	0.00125	0.00043	0.00019	0.00065	0.00024

Notes: Difference-in-differences estimates on the website-core subsample. *Restricted* in the Sample row indicates this restriction: the estimation sample comprises ventures covered by the BuiltWith data and active in the eleven markets in which the website is central to value creation (see Section 4). Dependent variables are the natural log of (one plus) the count of active tools: overall stack (Column 1), customer-input tools (Column 2), pre-launch tools (Column 3), A/B-testing and feature-flag tools (Column 4), and commerce tools (Column 5). Ever-carrier firms is the number of ventures ever carrying a tool of the given class. As discussed in the text, static coefficients are interpreted jointly with the event-study evidence in Figures 9 and 10. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

so the estimates speak to the thirteen months immediately following the launch rather than the full horizon; and filings are rare events (8,370 filing months in the estimation sample, 1,258 of them post-shock), so we report event counts throughout.

Figure 11 shows the event study for the any-filing margin, and it is the cleanest pre-trend pattern in the paper: pre-treatment coefficients are centered on zero throughout the pre-period with no systematic trend—though the earliest intervals are wide—and show none of the convergence dynamics that complicate other outcomes.

Figure 11: Dynamic DiD Event Study: Any Trademark Filing



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the any-filing indicator, with 95% confidence intervals. Effects relative to $t = -1$. Trademark coverage ends in December 2023, so the post-launch window extends only to $t = +12$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

Against that backdrop, the static estimate is marginally negative (-0.001, $p < 0.10$; Table 10). We do not read it as evidence of a decline, resting instead on the evidence around it: a flat event study, an observation window of only thirteen months, and a formal-registration margin that many software launches skip. The conclusion is scoped accordingly: it concerns trademark filings, not product introduction generally.

Taken together with Section 6.3, the product-development evidence is coherent: the reallocation away

Table 10: Difference-in-Differences: Trademark Filings

	(1) Any Filing
ChatGPT × AI-Ready	-0.001*
	(0.001)
Observations	1254466
Firm FE	Yes
Period FE	Yes
Controls	Yes
Post-shock window	13 months
Filing events	8,370
R-squared	0.00004

Notes: Difference-in-differences estimates. Dependent variable is an indicator for any trademark application filed in the month, with contemporaneous market-level controls included (see Table 5 notes). The sample is restricted to ventures covered by the trademark data and to the coverage window (January 2018 – December 2023), leaving thirteen post-shock months. Filing events is the number of firm-months with at least one filing in the estimation sample. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

from experimentation labor is not accompanied, at least in the first year, by any detectable change on the trademark-filing margin, while the commercialization infrastructure on which existing products are sold expands. The action, in other words, is in how ventures build and sell, not yet in how many products they introduce.

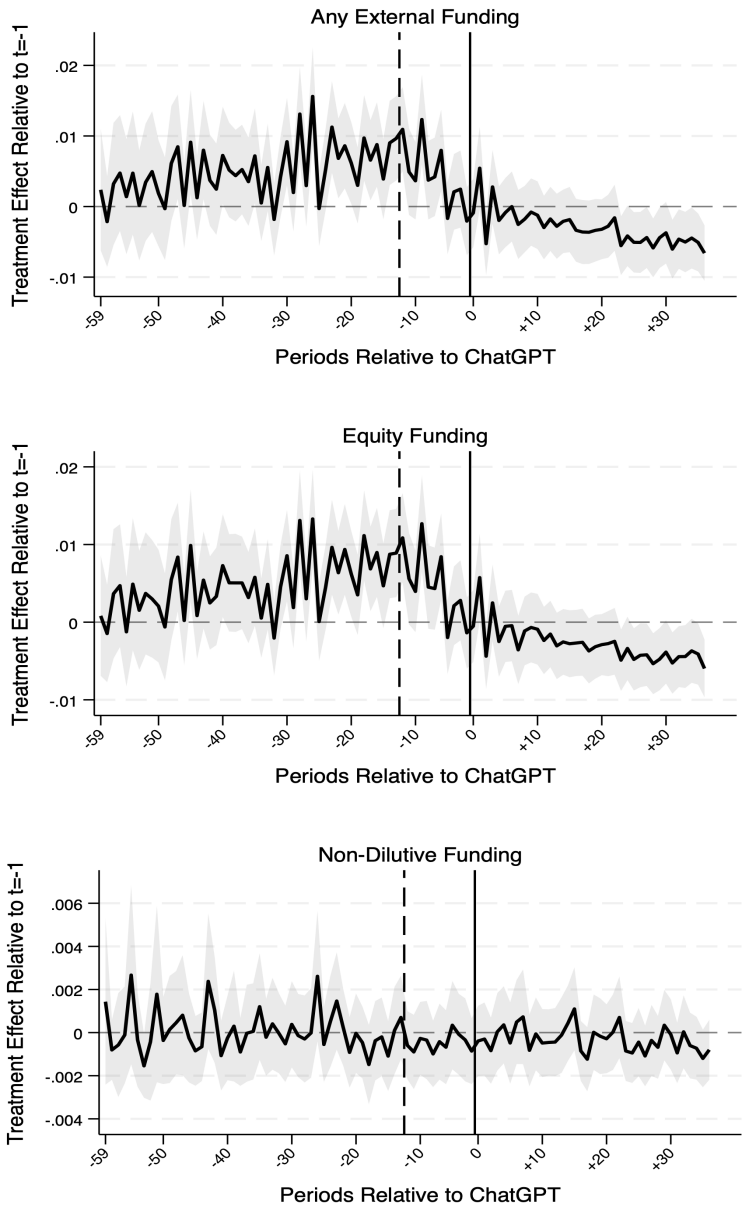
6.5 Financing and Exit

Finally, we ask whether the workforce reallocation is accompanied by changes in financing and performance. AI-ready ventures are, descriptively, prolific fundraisers: their differential funding-arrival rate runs roughly a percentage point above the reference-month gap at its pre-period peak (Figure 12). The question is whether the launch changed anything. We examine arrival rates for external, equity, and non-dilutive funding in turn.

The event study (Figure 12) shows why the static arrival decline reported below should not be read causally: treated ventures' differential funding-arrival advantage relative to the reference month (roughly one percentage point at its peak) erodes steadily from two years *before* the launch, and the post-period continues that slide without any visible break in November 2022. This is the funding winter meeting a population of high-raising growth ventures, not a response to ChatGPT.

Table 11 reports the arrival estimates: the monthly probability of raising any external round declines by 0.8 percentage points among AI-ready ventures post-launch ($p < 0.01$), with an identical estimate for equity rounds and a precise zero for non-dilutive funding. Given a base rate below one percent, the static

Figure 12: Dynamic DiD Event Studies: Funding Arrival



Notes: Coefficients $\hat{\beta}_\tau$ from the dynamic DiD specification for the three funding-arrival indicators (any external, equity, non-dilutive), with 95% confidence intervals. Effects relative to $t = -1$. The solid vertical line marks the ChatGPT launch (November 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). All regressions include firm and period fixed effects. Standard errors clustered at the firm level.

coefficient looks large, but the event-study evidence above shows the decline predates the launch.

Table 11: Difference-in-Differences: Funding Arrival by Type

	(1) Any External	(2) Equity	(3) Non-Dilutive
ChatGPT × AI-Ready	-0.008***	-0.008***	-0.000
	(0.000)	(0.000)	(0.000)
Observations	2064493	2064493	2064493
Model	LPM	LPM	LPM
Firm FE	Yes	Yes	Yes
Period FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Funding months	15,302	13,502	1,890
R-squared	0.00082	0.00091	0.00000

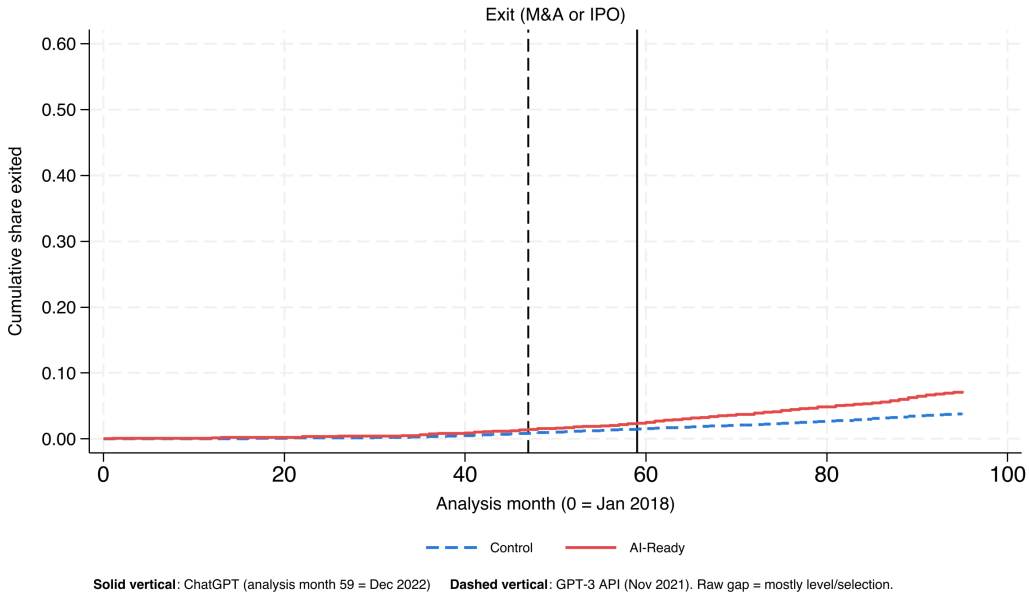
Notes: Difference-in-differences estimates. Dependent variables are monthly funding-arrival indicators: any external round (Column 1), any equity round (Column 2), and any non-dilutive round (Column 3). All columns include contemporaneous market-level controls (see Table 5 notes). Funding months is the number of firm-months with an event of the given type in the estimation sample. As discussed in the text, the event-study evidence in Figure 12 precludes a causal reading of the arrival decline. All regressions include firm and period fixed effects. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We conclude that financing shows no launch-timed break: the differential was already eroding before November 2022. Because the design cannot recover the counterfactual continuation of that pre-trend, we treat the financing margin as uninformative for the channels rather than as an estimated zero. The funding advantage itself is a matter of who these ventures are rather than what the shock did.

Figure 13 plots the Kaplan–Meier cumulative incidence of exit by treatment status. The two groups are close to indistinguishable over the first three years of the panel; a gap then opens and widens steadily through the post-launch window, with AI-ready ventures ending the panel at roughly seven percent cumulative exit incidence against roughly four percent for controls. The raw divergence mixes two things, the level difference in exit propensity (who these ventures are) and any post-launch response, and because it begins well before the launch, it cannot by itself establish a ChatGPT effect. The hazard models below separate the level from the post-launch differential.

The one performance margin that does respond is exit. We estimate Cox models for time to first exit (acquisition or IPO, whichever comes first; acquisitions account for the overwhelming majority of the 1,140 in-window events), with the shock entering as a time-varying covariate. Table 12 reports the results. The naive specification puts the post-launch exit hazard of AI-ready ventures at 2.1 times that of the baseline ($p < 0.01$), robust to stratifying the baseline hazard by founding cohort. Unlike the financing margin, where the event-study dynamics reveal the static decline to be a continuation of a pre-existing slide, the exit

Figure 13: Cumulative Exit Incidence by Treatment Status



Notes: Kaplan–Meier cumulative incidence of exit, defined as the earlier of the venture's first acquisition (excluding leveraged and management buyouts) and first IPO, by treatment status. Analysis time is months since January 2018, with ventures entering the risk set in their first panel month; exits are dated by announcement. The solid vertical line marks the ChatGPT launch (analysis month 59 = December 2022); the dashed vertical line marks the public release of the GPT-3 API (November 2021). The raw gap combines the AI-ready level difference with any post-launch response; the Cox estimates in Table 12 separate the two.

result survives the separation of the treatment-level effect: with the AI-ready level included, the post-launch differential remains at 1.41 ($p < 0.05$). AI-ready ventures were always more likely to exit; after ChatGPT's launch, they became more likely still, relative to their own elevated baseline.

We interpret this result cautiously but find it suggestive. The estimate is not a matched comparison; exits are dated by announcement; and we observe no transaction prices, so we cannot distinguish premium exits from distressed ones. Still, the post-2022 acquisition market makes acquirer demand for AI-capable teams and products one natural reading; whether the post-launch reallocation we document entered acquirers' valuations is a conjecture our data cannot test. An equally admissible reading, however, requires no link to the labor response at all: a post-2022 acquisition wave targeting AI-labeled firms generally would raise treated ventures' exit hazard mechanically, a treatment-correlated demand shock in the acquisition market, so the alignment with the reallocation story is suggestive rather than probative. On the first reading, the labor response documented in Sections 6.1 and 6.2 and the exit response documented here would be two sides of the same repricing, though we emphasize the reading is speculative: the estimate is a hazard differential, not a valuation result.

The exit result also bears on the precision channel, and the confrontation is worth making explicit: exit is a positive performance outcome, so its rise might seem to rescue precision. It does not. Precision predicts improvement through learning-driven product-market outcomes. Neither margin provides support: the trademark-filing margin shows no detectable change in its short window, and the financing margin is uninformative given its pre-trend. An exit hazard that rises unaccompanied by those improvements, in a post-2022 acquisition market actively seeking AI-capable teams, reads as a repricing of AI-readiness rather than as the harvest of sharper experimental learning.

Table 12: Time to Exit: Cox Proportional Hazard Estimates

	(1) Exit (M&A or IPO)	(2) Exit (M&A or IPO)	(3) Exit (M&A or IPO)
ChatGPT × AI-Ready	2.131***	2.138***	1.415**
	(0.169)	(0.169)	(0.199)
AI-Ready (level)			1.506***
			(0.176)
Observations	53120	53120	53120
Model	Cox PH	Cox PH	Cox PH
Strata	None	Cohort	None
Events	1,140	1,140	1,140

Notes: Cox proportional hazard estimates (hazard ratios) for time to first exit, defined as the earlier of the venture's first acquisition (excluding leveraged and management buyouts) and first IPO, dated by announcement. The ChatGPT shock enters as a time-varying covariate; risk time is censored at the end of the venture's panel life; recorded closure is thus treated as censoring, so estimates are cause-specific hazards of exit among ventures still operating, and the Kaplan–Meier complement in Figure 13 is descriptive rather than a competing-risks cumulative incidence; ventures enter the risk set in their first panel month. Observations are counting-process risk episodes: risk time is split at the ChatGPT launch, so ventures at risk in both periods contribute two episodes. The Cox specifications are estimated without the market-level controls used in the linear probability models, keeping the hazard specification parsimonious; the cohort stratification in Column 2 addresses composition. Column 2 stratifies the baseline hazard by founding cohort. Column 3 includes the AI-Ready level effect, so the interaction measures the post-shock differential relative to treated ventures' baseline exit propensity. Events is the number of in-window exits in the risk set. Standard errors clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

7 Discussion

We conclude by discussing the key findings, the contributions to research, and the limitations and open questions. Throughout, we distinguish what the evidence establishes causally from what it describes.

7.1 Key Findings

This paper examines how the emergence of generative AI reshapes experimentation in early-stage ventures. Our main finding is a specific compositional response. Following ChatGPT's launch, AI-ready ventures

(those whose top management teams possessed deep learning or machine learning expertise beforehand) reduced the share of their workforce allocated to experimentation by roughly 13% relative to its baseline, alongside a decline in the Engineering share of the team that grows to roughly two percentage points by the end of the window. The experimentation-share effect appears with the same sign under every treatment definition in our measurement grid, with significance in three of four cells, and both the experimentation and Engineering effects diverge following the launch after twenty months of flat differential pre-trends. The share of the workforce devoted to scaling activities is unaffected, indicating a response specific to experimentation rather than a shift in venture lifecycle posture.

The compositional shift is concentrated in prototyping (inductive, artifact-based market learning), while the shares allocated to customer input gathering and to formal controlled testing sit on precise zeros. This asymmetry is informative: the type of experimentation whose share declines is precisely the one involving generative, design-oriented tasks (building MVPs, iterating on prototypes, validating concepts), which are the activities most readily performed by AI-assisted generation and rapid prototyping tools. Customer research and formal analytical testing appear resilient, consistent with their value deriving from real-world behavioral data and statistical rigor that GenAI cannot yet credibly replicate.

AI-ready ventures' technology stacks expand after the launch, a level result we read descriptively, absent dynamic evidence (Section 6.3). Adoption of experimentation tooling shows no post-launch acceleration: paths are flat to mildly declining, the controlled-testing path drifts mildly negative in website-core markets, and pre-launch tooling is detection-limited. Commerce infrastructure (checkout, payment, and billing tooling) rises on flat recent pre-trends. Trademark filings show no detectable change within the thirteen-month coverage window, a null we rest on the flat event study, the short window, and a formal-registration margin many software launches skip. The funding advantage of AI-ready ventures is large, but its erosion predates the launch and shows no break at it. And the hazard of exit (acquisition or IPO) rises by roughly 40% net of the AI-ready level effect.

The precision channel receives no affirmative support on any margin where it makes a distinctive prediction, under the auxiliary assumption of Section 2 linking learning quality to observable product-market outcomes. Its distinctive footprint sits at the measurement step and in learning-driven outcomes: the controlled-testing share is a precise zero, the trademark-filing margin shows no detectable change in its short window, and the financing margin is uninformative given its pre-trend. The expansionary reading of the efficiency channel finds no support either: cheaper experimentation produced no complementary tooling

buildout. The two labor-saving readings, fixed-quantity efficiency and substitution, are the most likely, and they are observationally equivalent in these data, sharing the core fact that GenAI performs building labor: prototyping and engineering execution.

7.2 Contributions to Research

To our knowledge, this is among the first papers to examine how generative AI reshapes experimentation in new ventures, opening a research agenda at the intersection of entrepreneurship, organizational learning, and AI. We make three contributions: to research on GenAI and entrepreneurial learning, to absorptive capacity theory, and to the AI-and-jobs debate. We take each in turn.

Our first contribution is joint: to research on GenAI in organizations and to the entrepreneurship and organizational learning literatures on experimental learning. Prior work on GenAI has focused primarily on individual productivity, labor market outcomes, and new firm entry; we shift the lens to within-firm capability building, providing large-scale firm-level evidence on how a general-purpose technological shock restructures the internal learning processes of existing ventures. This margin matters because experimentation is a primary mechanism through which entrepreneurs resolve uncertainty about product, market, and business model viability, and the lean startup methodology that has dominated practitioner and academic thinking for over a decade centers on iterative field experimentation, prototyping in particular, as the core of that learning. Our evidence shows that GenAI's arrival reduced the share of venture workforces devoted to exactly that activity, concentrated in the prototyping roles where its generative capabilities operate, with no offsetting investment in experimentation tooling. The reallocation is compositional rather than absolute (AI-ready ventures remain large employers of experimentation talent), but the direction is unambiguous: at the margin, GenAI reduces the share of venture workforces devoted to building experiments. Whether that shift preserves the quality of entrepreneurial learning remains an open question.

Second, we contribute to absorptive capacity theory by providing large-scale quasi-experimental evidence that domain-specific prior knowledge conditions how firms respond to a general-purpose technology shock. The differential response we document—a post-launch workforce reallocation concentrated among ventures with ex ante DL/ML expertise—is the pattern the framework of Cohen and Levinthal (1990) predicts under our maintained exposure assumption; we observe the reallocation and the prior expertise, not adoption itself. Firms that have invested in a technical domain are better positioned to recognize, assimilate, and apply new knowledge in that domain when it becomes available; firms without such prior investment risk lockout

even when the technology is freely accessible. Our setting provides a clean and consequential test of this mechanism at the level of the venture.

Third, the evidence speaks to the debate on AI and jobs. That debate centers on displacement (whether AI destroys jobs and shrinks headcount), and the emerging empirical literature has likewise focused on entry, labor supply, and headcount margins (Bao et al., 2025; Yiu et al., 2026; Zhu et al., 2026). Our venture-level evidence shows that, inside firms, the margin we identify causally is task recomposition: GenAI hits specific task content (prototype-building and engineering execution) while employment descriptively grows, though our design is silent on the headcount counterfactual. The skill-conditioning result reinforces the point: DL/ML expertise gates the response, so GenAI's labor-market effects run through who can absorb the technology, not only through what it automates. For research on technology and labor, the venture sector thus offers an early view of a general pattern: the unit of impact is the task bundle within jobs, and absorptive capacity determines where the impact lands first.

7.3 Limitations and Open Questions

Several limitations of the current analysis deserve acknowledgment. Four matter most: imperfect measurement of the outcome, the properties of the underlying resume data, the maintained exposure assumption, and bounded generalizability. We discuss each in turn.

First, we measure the outcome imperfectly. We observe workforce allocation to experimentation-related roles rather than experimentation activity itself: we cannot see how many experiments AI-ready ventures ran, how they were designed, or what they revealed. This is precisely what prevents us from separating the two labor-saving channels: running the same experiments more cheaply and replacing field learning with synthetic feedback are observationally identical in workforce and tooling data. Distinguishing them requires direct measurement of experimental activity, for instance firm-level usage of generative-AI tools.

Second, the resume data carry their own measurement structure. Revelio positions are self-reported and updated with a lag, and a venture's employees enter the data only if they maintain an online profile; both the reporting lag and selection into profile coverage could in principle differ across treatment groups, and neither is directly observable. The reporting lag also interacts with the end of the panel window: positions begun in the final months may not yet be recorded, so late-window months are measured with more noise than earlier ones. These concerns are distinct from the classification issues of Section 4, and firm fixed effects

absorb only their time-invariant components.

Third, our exposure design requires the maintained assumption that DL/ML expertise in the TMT captures differential absorptive capacity with respect to GenAI. The assumption is theoretically grounded but cannot be directly tested, and AI-ready founders may differ on unobserved dimensions (risk tolerance, ambition, organizational culture) that independently affect both experimentation practices and post-shock hiring. We also pool DL and ML expertise into a single binary treatment, though DL-trained founders were almost certainly better positioned to exploit ChatGPT's capabilities than classical ML practitioners; Column 3 of Table 7 shows the result is unchanged when treatment requires DL expertise specifically, and a fuller DL-versus-ML decomposition remains a direction for future work.

Fourth, our sample is restricted to U.S. software ventures founded between 2018 and 2021, which limits the generalizability of our findings to other industries, geographies, and venture stages. Software ventures are arguably the most natural early adopters of GenAI tools, and the patterns we document may be more muted or differently structured in hardware, life sciences, or non-digital service contexts.

These limitations chart the agenda ahead. The immediate task is to observe experimentation directly—firm-level usage of generative-AI tools, and the number and design of the experiments ventures run—so as to separate the two labor-saving readings our data cannot distinguish. The deeper task is to connect the reallocation to learning outcomes: whether ventures that devote less labor to building experiments learn as well as those that build by hand. On the answer turns whether the pattern we document is a productivity gain in entrepreneurial learning or a corner quietly cut.

References

Agrawal, A., Camuffo, A., Gambardella, A., Gans, J., Scott, E., & Stern, S. (2025). *Bayesian entrepreneurship: A framework for decision-making in uncertain environments*. MIT Press.

Altman, S. (2022). ChatGPT five days in. https://twitter.com/sama/status/1599668808285028353

Babina, T., Fedyk, A., He, A. X., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, *151*. https://doi.org/10.1016/j.jfineco.2023.103745

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

Blank, S. (2013). Why the Lean Start-Up Changes Everything. *Harvard Business Review*, *91*, 63–72.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*.

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). *Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics* (Working Paper No. 24001). National Bureau of Economic Research. https://doi.org/10.3386/w24001

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A Scientific Approach to Entrepreneurial Decision-Making: Evidence from a Randomized Control Trial. *Management Science*, *66*(2), 564–586.

Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., & Spina, C. (2024). A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*.

Chavda, A., Gans, J. S., & Stern, S. (2024). Theory-based entrepreneurial search. *Strategy Science*, *9*(4), 397–415.

Chen, A. (2025, September 3). *AI will change how we build startups—but how?* https://andrewchen.com/ai-will-change-how-we-build-startups-but-how/

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, *35*(1), 128–152.

Contigiani, A., Denoo, L., & Hablicsek, M. (2026). Toward a theory of bayesian experimentation in early-stage ventures. *Strategic Entrepreneurship Journal*.

Contigiani, A., & Levinthal, D. A. (2019). Situating the construct of lean start-up: adjacent conversations and possible future directions. *Industrial and Corporate Change*, *28*(3), 551–564.

Eisenhardt, K. M., & Bingham, C. B. (2017). Superior Strategy in Entrepreneurial Settings: Thinking, Doing, and the Logic of Opportunity. *Strategy Science*, *2*(4), 246–257.

Eisfeldt, A. L., & Schubert, G. (2025). Generative AI and finance. *Annual Review of Financial Economics*, *17*.

Gofman, M., & Jin, Z. (2024). Artificial intelligence, education, and entrepreneurship. *The Journal of Finance*, *79*(1), 631–667.

Impink, S. M. (2025). General-purpose technologies and differentiation: Evidence from startup cloud adoption [Available at SSRN 4200675].

Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as Experimentation. *Journal of Economic Perspectives*, *28*(3), 25–48.

Koning, R., Hasan, S., & Chatterji, A. (2022). Experimentation and start-up performance: Evidence from A/B testing. *Management Science*, *68*(9), 6434–6453.

Marchesi, E., & Tang, S. (2025). *When AI meets entrepreneurship: Evidence from the commercialization of ChatGPT* [Available at SSRN 5759264].

OpenAI. (2019). Better language models and their implications.

OpenAI. (2020). GPT-3 API: General availability.

OpenAI. (2021). OpenAI codex.

OpenAI. (2022). ChatGPT: Optimizing language models for dialogue.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, *35*.

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving language understanding by generative pre-training* (tech. rep.). OpenAI. https://openai.com/research/language-unsupervised

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language models are unsupervised multitask learners* (tech. rep.). OpenAI. https://openai.com/research/better-language-models

Reynolds, J. (2025). ChatGPT has more than 800 million weekly active users, says Altman. https://tech.eu/2025/10/07/chatgpt-has-more-than-800m-weekly-active-users-says-altman/

Ries, E. (2011). *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. Crown Business.

Ries, E. (2017). *The Startup Way: How Modern Companies Use Entrepreneurial Management to Transform Culture and Drive Long-Term Growth*. Currency.

Stroube, B., & Dushnitsky, G. (2023). *Built with BuiltWith: A technical note* [Available at SSRN 4475309].

Stroube, B., & Dushnitsky, G. (2025). *Web technology data for strategy and entrepreneurship research* [Working paper, London Business School].

UBS. (2023). ChatGPT: The fastest growing app in history.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Wang, X., & Wu, L. (2026). Artificial intelligence, lean startup method, and product innovations. *Management Science*, *72*(1), 756–782.

Wiggers, K. (2022, October). Revelio labs raises cash to scrape the public web for HR insights [October 6, 2022]. https://techcrunch.com/2022/10/06/revelio-labs-raises-cash-to-scrape-the-public-web-for-hr-insights/

Wright, N. L., & Saiedi, E. (2025). The mid-sized market trap in entrepreneurial scaling [Available at SSRN 5078041].

Yerramilli-Rao, B., Corwin, J., Li, Y., & Lakhani, K. R. (2025). Strategy in an era of abundant expertise. *Harvard Business Review*, *103*(2), 72–81.

Yiu, S., Seamans, R., Raj, M., & Liu, T. (2026). Strategic responses to technological change: Evidence from online labor markets. *Organization Science*.

Zhu, R., Knight, S., Colicev, A., & Bart, Y. (2026). The effect of AI on marketing employment: Evidence from 110m+ online employment records. *Available at SSRN 6653422*.