Mapping AI into Production: A Field Experiment on Firm Performance*

Hyunjin Kim Dahyeon Kim Rembrand Koning

March 20, 2026

AI can deliver productivity gains on individual tasks, yet evidence on whether these gains aggregate to firm performance remains limited. We study a central friction in AI adoption, which we call the mapping problem: discovering where and how AI creates value within a firm's production process. Across 515 high-growth startups, we run a field experiment in which treated firms receive information about how other firms have reorganized production around AI, prompting them to search for use cases across a broader set of firm functions. We find that treated firms discover more AI use cases, a 44% increase, concentrated in product development and strategy. These changes result in economically meaningful performance gains. Treated firms complete 12% more tasks, are 18% more likely to acquire paying customers, and generate 1.9x higher revenue. Revenue and investment gains are largest at the 90th percentile and above, consistent with AI expanding the upper range of what firms achieve rather than modestly improving marginal ventures. Despite faster growth, treated firms do not scale inputs proportionally. Their demand for external capital investment falls by just over $220,000, a 39.5% decrease, relative to the control group, while their demand for labor remains unchanged. These results provide causal evidence that AI improves firm performance and productivity even at its current capabilities, and that discovering where and how to deploy AI is a key bottleneck in realizing the gains from this technology.

*H. Kim (corresponding author): h@insead.edu. D. Kim: dahyeon.kim@insead.edu. R. Koning: rem@hbs.edu. The authors gratefully acknowledge financial support from INSEAD R&D and the INSEAD AI Venture Lab. The authors thank Aasiya Ahmed, Sadig Huseynzade, Ashish Kulkarni, Simonne Pinto, Bingrui Zhong for excellent research assistance. We thank Ajay Agrawal, Rebecca Henderson, Phanish Puranam, Michelle Rogan, and Scott Stern, along with seminar participants at MIT Sloan, ESSEC, LBS, UCL, University of Oxford, and the INSEAD brownbag for their helpful feedback. The paper was formerly entitled "Reimagining the Firm with AI: Experimental Evidence from Startups." This experiment was approved by the INSEAD Institutional Review Board and was preregistered on the AEA RCT Registry under trial number AEARCTR-0016746.

1 Introduction

A central promise of AI rests in the idea that it will make firms more productive and drive economic growth. The breadth of activities that AI can reshape within the firm—from operations (Brynjolfsson, Li, and Raymond, 2023) to strategic decisions (Csaszar, Ketkar, and Kim, 2024)—makes this promise especially compelling. While much task-level evidence shows productivity gains across a wide range of firm activities (Noy and Zhang, 2023; Dell'Acqua et al., 2023), these gains have largely failed to aggregate to firm-level performance, echoing a pattern from earlier general-purpose technologies (David, 1990; Brynjolfsson, Rock, and Syverson, 2018). Theoretical models of the impact of AI on organizations highlight the challenges of improving firm-level outcomes (Brynjolfsson, Rock, and Syverson, 2021; Gans and Goldfarb, 2026). Compounding these theoretical considerations, the rapid adoption of AI and the wide variation in tools used across firms has made it empirically difficult to isolate how AI affects firm performance (Seamans and Raj, 2018).

In this paper, we estimate the causal effect of AI use on firm performance by leveraging a key friction, which we call the *mapping problem*: the challenge of discovering where and how AI creates value within a firm's production process. Prior general-purpose technologies reshaped concrete processes—electricity demanded new factory layouts, the internet required new distribution channels—and a large literature has studied the complementary investments needed to restructure production around them (David, 1990; Milgrom and Roberts, 1990; Bresnahan et al., 1996). With AI, the space of possible applications is especially vast and flexible: the same technology can assist with writing an email or fundamentally change how a product is developed, yet even closely related tasks can differ sharply in how well AI performs (Dell'Acqua et al., 2023; Vafa, Rambachan, and Mullainathan, 2024; Otis et al., 2024). As a result, a central friction is the *discovery* of how to reorganize, not just the investments in reorganization itself. Discovering where and how AI creates value is fundamentally a search problem, and a large strategy literature suggests that search is often local (Levinthal and March, 1993; Stuart and Podolny, 1996; Kim and Wu, 2025), especially when confronted with jagged, combinatorial, and vast landscapes like those introduced by AI (Fleming, 2001). This suggests that firms may default to obvious local applications, while higher-value uses remain undiscovered.

We leverage this mapping friction to design a treatment that induces exogenous variation in firms' ability to discover how to use AI in production. In a 3-month accelerator with 515 high-growth startups from across the globe, we randomize the information provided through the accelerator's

workshops: treated firms received case studies showing how other firms have reorganized production around AI. These case studies included discussions of how AI-native firms have reorganized their workflows, teams, business models, and financing to maximize the benefits of using AI (Koning, Keller, and Kim, 2025; Koning, Kim, and Keller, 2025). Beyond these workshops, treated and control ventures all received standard accelerator support (Cohen et al., 2019), including general entrepreneurship case studies, technical AI training, API credits, mentorship, and investor exposure. As a result, our design lets us both estimate the causal impact of AI on firm-level outcomes and isolate the role of mapping frictions in explaining why AI's firm-level impact has lagged its capabilities.

Our treatment leads to changes in how firms use AI, which in turn produce economically meaningful performance gains. Treated firms discover 2.7 additional AI use cases (a 44% increase), which span a broader set of activities across the firm and are especially concentrated in product development and strategy-related domains. These changes in AI use lead to measurable gains in performance: treated firms complete 12% more tasks, are 11 percentage points (18%) more likely to acquire paying customers, and ultimately generate 1.9x higher revenues compared to control firms. Revenue and investment gains are largest at the 90th percentile and above, consistent with AI helping founders unlock bottlenecks that constrain the upper range of venture performance. These gains are broad-based: we find no significant differential effects by founder technical background or baseline firm performance, consistent with the mapping problem binding across a wide range of firms and founders. Despite faster growth, treated firms do not scale inputs proportionally: their demand for external investment falls by just over $220,000 relative to the control group, while their demand for labor remains unchanged, consistent with the idea that AI enables firms to produce more output with the same resources.

Our design also allows us to rule out a range of non-AI related explanations and to estimate the returns to AI use within the firm. Attrition from the accelerator was low (1.6%, eight ventures) and balanced across treatment and control, suggesting similar engagement across groups. Analysis of participant feedback shows that control and treated founders expressed comparable sentiment about the accelerator experience, further reducing concerns about motivation-driven effects. Moreover, differences between control and treated firms are driven by changes in AI use cases, making it unlikely that the results reflect non-AI channels. This pattern is consistent with the nature of the intervention: the treatment helped founders explore how to reorganize their ventures so that they could benefit from AI, rather than provide additional general management or entrepreneurship

training (Bloom et al., 2013; Camuffo et al., 2020; Dimitriadis and Koning, 2022). Instrumenting AI use cases with treatment assignment reinforces this interpretation. Our IV estimates suggest that each additional AI use case prompted by treatment leads to 0.85 more completed tasks and approximately 26% higher revenue. These are large effects, suggesting that AI is fundamentally reshaping how ventures scale when they can map it across its production process.

Our results make three contributions. First, we provide causal evidence that AI's task-level productivity gains can aggregate to firm-level performance — and identify a key friction that explains why they often do not. Despite widespread evidence documenting productivity gains at the task level (Brynjolfsson, Li, and Raymond, 2023; Noy and Zhang, 2023; Dell'Acqua et al., 2023), the question of whether these aggregate to firm-level outcomes has remained open—theoretically, because complementarities and intangible investments can offset or delay returns (Gans and Goldfarb, 2026; Brynjolfsson, Rock, and Syverson, 2021), and empirically, because existing firm-level evidence is largely correlational (Babina et al., 2024; **?**; **?**). We show that AI's benefits do aggregate to the firm level, with gains that are broad-based across a wide range of firms.

Second, we show that the binding constraint on AI's firm-level value is not access to the technology but search over where to deploy it. We call this the mapping problem. Most existing evidence on the impact of AI comes from variation in *whether* firms adopt AI at all. We instead generate variation in *how* firms map AI into their production processes. Our treatment reveals that firms tend to search too narrowly: without examples or structured frameworks that broaden the space they consider (Kim and Wu, 2025), firms default to familiar applications and miss higher-value uses embedded in their production processes, consistent with a large literature in strategy on boundedly rational search (Gavetti and Levinthal, 2000; Levinthal and March, 1993). This idea also sheds light on the role of complementarities in the adoption of general purpose technologies such as AI (Bresnahan et al., 1996; Brynjolfsson, Rock, and Syverson, 2021): before firms can invest in redesigning their processes, they must first discover *where* the complementarities lie.

Third, we show that AI is reshaping how firms are created and the scale they can reach. Treated firms achieve faster growth without proportional increases in labor or capital, consistent with a reduction in the costs of experimentation and scaling seen in earlier technological waves (Dushnitsky and Stroube, 2021; Ewens, Nanda, and Rhodes-Kropf, 2018). But our upper-tail results are harder to explain by cost reduction alone, and are instead consistent with complementarities across activities within the firm amplifying the returns to AI (Kremer, 1993; Gans and Goldfarb, 2026). These patterns suggest that AI is not just lowering the costs of entry or modestly improving middling

ideas (Reimers and Waldfogel, 2026), but expanding the frontier of what firms can be built.

2 The Mapping Problem

A growing body of task-level experiments shows that AI produces meaningful productivity gains across a wide range of work settings (Brynjolfsson, Li, and Raymond, 2023; Noy and Zhang, 2023; Dell'Acqua et al., 2023). Yet, aggregating these gains to the firm level is far from automatic. A firm's output depends on the joint quality of many interconnected, complementary activities (Kremer, 1993; Gans and Goldfarb, 2026), and a large literature on general-purpose technologies shows that realizing their benefits requires reorganizing production around them — costly investments in process redesign, restructuring, and new skills that take time to manifest (David, 1990; Milgrom and Roberts, 1990; Bresnahan et al., 1996). Without these investments, a task-level improvement may be bottlenecked by unchanged processes elsewhere in the firm, producing limited or no gains in overall performance (Brynjolfsson, Rock, and Syverson, 2021). Realizing the returns to AI thus requires not just adopting the technology but reorganizing production around it.

This complementary structure implies that before firms make these investments to benefit from AI, managers must solve a search problem: they must discover which activities within their firm's production process AI can improve and how to adjust the rest of the firm. We call this the *mapping problem*.

The mapping problem is difficult for three reasons. First, AI's capabilities are uneven and hard to predict. Even closely related tasks can differ sharply in how well AI performs them, and people—including experts—systematically misjudge where AI will help (Dell'Acqua et al., 2023; Vafa, Rambachan, and Mullainathan, 2024; Otis et al., 2024). This makes it hard for managers to assess, ex ante, which parts of their production process would benefit from AI. Second, the search space is vast. Within any given firm, AI could potentially be applied to dozens of activities. A large literature in strategy shows that managers search locally in the neighborhood of what they already know (Gavetti and Levinthal, 2000; Levinthal and March, 1993), absent scaffolding like analogical reasoning and structured frameworks (Gavetti, Levinthal, and Rivkin, 2005; Kim and Wu, 2025). Applied to AI, this means firms default to obvious applications—a chatbot for customer support, an LLM to draft emails—while the highest-value applications that require rethinking how work is organized are likely to go undiscovered. Third, complementarities compound the difficulty. When activities within the firm are tightly coupled, the value of applying AI to one activity depends on

whether adjacent activities also adjust (Siggelkow, 2002; Bresnahan et al., 1996). This means that partial automation can preserve bottlenecks rather than relieve them (Demirer et al., 2025).

Consider a firm that uses AI to accelerate product development—generating prototypes, writing code, testing features faster than before. If the firm does not also rethink how it identifies what customers actually want, it simply builds the wrong product faster. The bottleneck has moved, not disappeared. The returns to AI depend not on any single application but on whether the firm reorganizes the surrounding work to take advantage of it. When it does—rethinking its full production process and discovering where AI can be applied to improve not just one activity but unlock value in adjacent ones—the gains compound. When it does not, task-level productivity gains dissipate before they reach the firm's bottom line.

Importantly, the mapping problem is a friction in discovery, not in access. Two firms with identical tools, training, and budgets can realize very different returns if one searches more broadly across its production process for where the technology creates value. This distinguishes the mapping problem from the access and skill frictions that dominate both the current policy conversation and the existing empirical literature on technology adoption. Providing firms with a new technology and teaching them how to use it may be insufficient; what matters is whether they can discover where in their production process to deploy it and how to organize the rest of the firm to benefit from it. While this friction is general—it arises whenever a technology's potential applications are vast, uneven, and hard to predict ex ante—it is especially acute for AI, where the same model can draft an email or redesign a product and even experts cannot reliably distinguish which use will create more value ex ante.

We study the mapping problem in the context of early-stage startups in an accelerator setting, which offers an unusually clean context to isolate this friction. Their production processes span multiple functions, organizational inertia is low, and the accelerator structure provides both a natural vehicle for delivering treatment and a rich, high-frequency window into how ventures apply AI across their operations, what tasks they complete, and how these changes aggregate to firm-level performance. We describe this setting next.

3 Empirical Context

We designed and ran our field experiment within the AI Founder Sprint, a three-month global, virtual startup accelerator at INSEAD designed to support early-stage ventures in building their

companies with AI. The accelerator was structured around six components—four shared equally across all firms, and two whose content varied between treatment and control. Table 1 provides an overview of the program and timeline.

Four components were shared by all firms. First, *resources*: participants received API credits, access to frontier AI models, and onboarding sessions from partners including Google Cloud, OpenAI, NVIDIA, and Manus AI, totaling approximately $25,000 in-kind per firm. Second, *technical training*: participants attended weekly three-hour hands-on sessions delivered by the MIT/Harvard Sundai AI Club covering rapid prototyping, context engineering, RAG, agentic workflows, and vibe-coding. Third, *demo days and pitch practice*: all firms had opportunities to pitch to leading venture capitalists from across the globe during the Sprint's VC Week, and were equally eligible to compete for $100,000 USD in non-dilutive seed funding at Demo Days in Singapore and Abu Dhabi. Fourth, all firms submitted weekly progress reports describing their activities, financial progress, and new AI use cases. These reports helped ventures stay on track and helped the accelerator team identify and respond to each firm's evolving needs. They also provided a rich source of data on venture traction and activities, which we describe in detail below.

Starting in the third week of the accelerator, two components varied between the treatment and control groups. First, *workshops*: all firms were required to attend a weekly 60-minute faculty-led session on Zoom that delivered a case-based entrepreneurship curriculum focused on standard best practices for making progress on early-stage venture ideas. Alongside these required workshops, all firms were invited to six optional 90-minute masterclass workshops with leading entrepreneurs, investors, and AI researchers.[1] As discussed below, starting in the third week, treated entrepreneurs received different information in their required workshops and were eligible for additional optional masterclass workshops.

The final component, *peer learning and office hours*, was available to all firms. All ventures had required weekly peer group meetings comprising four to eight founders with a venture guide, where they shared progress, challenges, and plans. They could all also sign up for one-on-one or small-group office hours with experienced mentors and accelerator instructors. These sessions were intended to help translate information provided in the workshops into firm-specific action. Since these sessions enabled participants to discuss and work together to implement the workshops' content, and since the workshop content varied across treatment and control, the peer learning

[1] Speakers included Ethan Mollick (Wharton), Grant Lee (Gamma), Hian Goh (Openspace Capital), Gaurav Jain (Afore Capital), Bill Tai (Angel Investor), Claudia Zeisberger (INSEAD), and Chris Yeh (Blitzscaling).

sessions naturally reinforced any differences between the two groups.

This setting provided an ideal context to study frictions in mapping AI to firm activities and processes. Early-stage ventures face relatively low organizational inertia, so difficulty in identifying and integrating high-value AI use cases in this context is informative—and, if anything, likely understates the magnitude of these frictions in larger firms. The accelerator context also provided unusually rich measurement: we observe a large sample of ventures and can track, week by week, how they apply AI across firm functions, what tasks they complete, and multiple dimensions of traction and performance. Finally, the program's global, online structure supports clean experimental variation with limited opportunities for spillover effects.

3.1 Sample selection

Ventures were selected through an open application process: any promising pre-seed venture could apply, regardless of affiliation with INSEAD. Applications collected information on founder characteristics, venture traction, and current AI usage. From the selected ventures, we restricted the experimental sample to those who provided research consent and completed baseline data collection in the first and second weeks of the accelerator, yielding a sample of 515 firms. Because treatment did not begin until the third week, all sample restrictions are based on pre-treatment data.

The median firm was founded in 2024 with a team of four. The sample spanned the globe, as shown in the map displayed in Figure A1: 33.8% of ventures were in the Middle East and Africa, 23.6% in Asia-Pacific, 22.3% in Europe, and 20.2% in the Americas. Many firms had achieved initial traction at the time of application: 52.6% had launched a product, 55.3% had acquired customers, 34.6% had generated revenue, and 35.3% had raised external investment. Among revenue-generating firms, median monthly revenue was $1,000; among funded ventures, the median investment raised was $10,000.

3.2 Experimental Design

3.2.1 Randomization

We implemented stratified randomization to assign enrolled firms to control and treatment conditions, stratifying on three pre-treatment characteristics: continent, traction score, and baseline AI use. Continent was defined based on firm location and categorized into four regions: Middle East and Africa, the Americas, Asia-Pacific, and Europe. Traction score was defined as the sum of three

binary indicators capturing whether the firm had launched a product, generated revenue, and raised external investment, yielding values ranging from 0 to 3. Baseline AI use was defined as a binary indicator based on the number of self-reported AI use cases at baseline: firms reporting 0-2 use cases (below the sample median) were classified as low AI use, and those reporting three or more use cases (at or above the median) were classified as high AI use.

These variables produced 32 strata, which had a minimum of six firms per stratum. Within each stratum, firms were randomly assigned to treatment or control, resulting in 255 treatment firms and 260 control firms. As shown in Table 2, baseline characteristics are well balanced across both groups.

3.2.2 Treatment Intervention: Mapping AI into Production

Identifying where AI will perform well is difficult even for experts (Dell'Acqua et al., 2023; Vafa, Rambachan, and Mullainathan, 2024), and understanding how to change firm-specific production processes to benefit from AI is harder still (Brynjolfsson, Rock, and Syverson, 2021; Kim et al., 2024; Otis et al., 2024). The complementary nature of firm activities compounds this difficulty: the firm-level effect of an improvement on an individual task depends on the network of other tasks that it is connected with (Kremer, 1993; Gans and Goldfarb, 2026). The returns to AI adoption at the firm level therefore depend not just on learning to use a particular AI tool, but on discovering where to use AI and how the rest of the firm must adjust to benefit from the technology (Bresnahan et al., 1996; Brynjolfsson, Rock, and Syverson, 2021).

Our intervention was designed to overcome these challenges by providing treated firms with information on how other firms have reorganized around AI. To deliver this information, we leveraged the accelerator's workshops as a setting where we could shift what treated firms learned relative to control. During both the required weekly workshops and several of the optional guest speaker sessions, treated firms received case studies focused on AI-native firms illustrating how they are reorganizing around this technology.[2] Crucially, these case studies went beyond which AI tools to use. They also illustrated where AI can be effectively applied and how the rest of the firm must change to make such applications effective. These discussions raised questions like whether AI demands different kinds of talent, how it changes the cadence of product development and operational processes, and whether it opens the door to fundamentally different business models.

[2]Control founders also engaged in weekly case discussions, but these focused on general entrepreneurship topics such as building an ideal customer profile, designing tests to validate ideas, and other common themes drawn from the lean startup and scientific approach to entrepreneurship literatures (Camuffo et al., 2020).

While these conversations were wide-ranging—they had to be, because best practices are still being established in this space—they shared a common thread of helping ventures think about how AI reshapes not just a single task but the structure of production in their firms.

To illustrate the content that treated firms received, Figure 1–Figure 4 present four diagrams corresponding to four core case discussions. Drawing on recent work that explores the impact of AI on chains of tasks within production processes (e.g., Demirer et al., 2025), each diagram uses a simple visual structure to show how production was reorganized relative to a version without AI. In each diagram, tasks are arranged in sequence with the responsible person (or AI) labeled beneath; Panel A shows the production sequence before AI and Panel B how it changes with AI.

The first two cases focus on product development, a central activity for any startup. Figure 1 presents Gamma, an AI-native presentations company that has grown to over $50M in annual recurring revenue with roughly 50 employees (Koning, Keller, and Kim, 2025). As Panel A illustrates, conventional product development requires a chain of specialists, each handling one step in cycles that can take months. With AI (Panel B), the chain compresses: AI detects usage patterns and generates product variants directly, enabling a single PM to continuously ship features that would previously have required an entire team. The reorganization also introduces a new task, developing AI evals, to ensure the rapidly generated variants actually improve the product. The key insight is that the gains come not from adopting AI for a single step but from deploying it across multiple steps and rethinking what that combination makes possible.

Figure 2 draws on a session led by Jordan Metzner of Ryz Labs on how AI changes prototyping. As Panel A shows, conventional prototyping is a sequential commitment: choose one tech stack, hire one engineering team, build one version often over months, then test. With AI (Panel B), the founder writes a PRD and feeds it into multiple AI coding tools simultaneously, building the same idea multiple ways rather than betting on a single approach. This creates a new bottleneck: with building no longer the constraint, the limit becomes how quickly the founder can get customer feedback. Here too, AI helps: because prototypes update in minutes, the founder can test the best version with a customer, build a feature live on the call, and test again.

The remaining two cases illustrate how the mapping problem extends well beyond product development to workflows, operations, and business models. Figure 3 presents FazeShift, an AI accounts receivable startup whose co-founder Caitlin Leksana led a session on workflow automation (Koning, Kim, and Keller, 2025). Panel A shows the conventional accounts receivable process: eight steps alternating between software systems and humans who bridge between them. With AI

(Panel B), this bridging is replaced by a fully automated sequence. The key insight is that the full chain of steps needs to be automated; partial automation preserves the bottleneck, preventing the value from accruing to customers (Demirer et al., 2025).

Figure 4 presents a case on Ranger, illustrating how reorganizing around AI can shift the business model itself. As Panel A shows, the conventional VC-funded path front-loads capital: the chain begins with "Raise money" and selling does not happen until step four. Panel B reorders the entire chain. The founder starts by selling QA testing as a service from day one, building tacit knowledge about which steps are routine and where AI can automate. By using AI to improve their own efficiency and that of future employees, the founder transforms the margins of the business from a services firm to something closer to a product firm. The most visible shift is that "Raise money" moves from the first box in Panel A to the last in Panel B, giving the venture optionality on whether and when to raise outside capital.

Across all four cases, a common theme emerged: the gains from AI depend not just on automating individual tasks but on reorganizing complementary activities so that improvements in one step are not bottlenecked by constraints in the next.

Peer learning and office hours reinforced the workshop content. As described above, all firms had required weekly peer group meetings and optional office hours with mentors and accelerator instructors. Because the workshop content differed across treatment and control, and peer sessions enabled participants to discuss and implement what they learned, these sessions naturally amplified treatment-control differences.

Finally, to minimize potential spillovers, treatment and control participants were placed into separate groups on the program's online platform and were unable to directly message or interact with members of the other group. During shared program components, interactive features including the Zoom chat function were disabled to prevent cross-group communication. These measures limited opportunities for participants to exchange information about workshop content or access materials outside their assigned condition.

4 Data and Empirical Specification

4.1 Data

We collected a wide range of data throughout the accelerator, including both quantitative measures of venture development and qualitative evidence of founders' behavior and thinking as reflected in

text data.

Baseline application materials. Prior to the accelerator, all ventures submitted application materials, including a description of their business, a pitch video, self-reported growth metrics, and additional information about the team. These data provide the basis for our pre-treatment covariates, enabling stratified randomization and the assessment of baseline balance across treatment arms.

Weekly progress reports. Our primary data source is a series of structured 'weekly progress reports' submitted by each firm throughout the Sprint. Each week, ventures completed a survey that collected data on venture progress, which reported: (1) growth and financial metrics including the number of customers, revenues, costs, and investments; (2) their anticipated demand for capital and labor by the Demo Days in December (over 3 months away from the start of the program); (3) up-to-date descriptions of their ventures and whether they had pivoted; (4) any new AI use cases in the last week; (5) specific tasks or actions they took to advance the startup that week; (6) product demos showing the changes they made; (7) in two of the weeks, pitch decks and videos. This weekly cadence of detailed data collection allows us to observe venture progression over time and construct cumulative measures of activity over the post-baseline period, in addition to endline snapshots of venture performance. Compliance was high: 98.4% of ventures (507 of 515) completed at least one post-baseline weekly progress report, and 81.0% completed all eight progress reports (Table A1).

Process & engagement data. We also collected supplementary data on venture engagement throughout the program. This includes recordings and transcripts of weekly guided peer group meetings, workshop attendance records, and discussion logs from our online community forum. These sources are used primarily for compliance checks and mechanism analyses rather than as primary outcome measures.

4.2 Outcome Measures

Our primary measure of AI usage is the number of distinct AI use cases reported by each firm, cumulated over the post-baseline period. Each week, founders described in their own words any new AI use cases they had implemented or explored. All responses were independently reviewed by two independent human coders, who assessed whether each entry constituted a valid AI application and

flagged duplicates (i.e., entries already reported by the same firm in a previous week). Disagreements were resolved by a third independent coder.[3] Full coding guidelines are provided in Appendix 7.2.1.

We also measure the pace and direction of venture activity. Each week, ventures reported the concrete tasks they completed to advance the startup. The number of concrete tasks completed is constructed analogously to AI use cases — cumulated over the post-baseline period, coded by two independent human coders and resolved by a third, and classified as either internal (e.g., product development, coding) or external (e.g., customer meetings, investor pitches) (Appendix 7.2.2).

We study a number of measures of venture performance. First, we construct an index of venture progress, which represents the sum of four binary indicators: whether the venture had launched its product, acquired customers, generated revenue, and raised investment at endline. We standardize this sum using the full-sample mean and standard deviation so that treatment effect estimates are interpretable in standard deviations. We also examine each of the four binary variables separately. In addition to these progress metrics, we measure the total amount of revenue generated, team size and composition, and investment raised for each venture by endline, our last week of data collection. We also track whether ventures pivoted from their original idea at any point during the Sprint and whether they were shut down at endline.

Finally, we measure ventures' estimated demand for inputs. We measure capital demand as the amount ventures reported they would seek to raise on Demo Day, and labor demand from founders' responses describing roles they would like to post job postings for. Each listed role was extracted from free-text, screened to exclude any invalid entries unrelated to hiring, and matched against a keyword list to assign a functional category, with tech-related postings defined as engineering, AI specialist, product, or design roles. We categorize job postings and examine differences in demand for tech-related roles to assess whether treatment meaningfully shifts the types of roles founders are seeking to hire.

4.3 Empirical specification

Our main specification estimates the intent-to-treat (ITT) effect of assignment to treatment:

$$y_i = \beta_1 \, Treatment_i + \delta_{s(i)} + \varepsilon_i \tag{1}$$

where y_i is the outcome of interest for firm i, $Treatment_i$ is an indicator for treatment assign-

[3] All coders were blind to the experiment, hypotheses, and treatment conditions.

ment, and $\delta_{s(i)}$ are fixed effects for randomization strata defined by continent, baseline traction, and baseline AI use. Standard errors are clustered at the firm level, the unit of randomization.

Heterogeneity Analysis. To examine heterogeneous treatment effects, we estimate specifications of the form:

$$y_i = \beta_2 \, Treatment_i + \gamma \, X_i + \phi \, (Treatment_i \times X_i) + \delta_{s(i)} + \varepsilon_i \tag{2}$$

where X_i is a pre-treatment firm or founder characteristic capturing treatment effect heterogeneity. The coefficient ϕ captures differential treatment effects as a function of X_i. We examine heterogeneity across pre-registered stratification dimensions such as baseline traction and baseline AI use, and also conduct exploratory analyses on additional founder and firm characteristics such as technical background, prior startup experience, and team size.

Instrumental Variables Specification. Our ITT estimates capture the effect of being assigned to treatment, but since treatment shifts the number of AI use cases firms adopt, we can go further and estimate the returns to each additional use case. This matters for two reasons. First, it sheds light on mechanism: if treatment effects are mediated by changes in AI use, channels unrelated to AI are unlikely to drive our results. Second, it provides a better estimate of AI's returns, since control firms also use AI—the ITT captures only the marginal difference in use cases between groups, not the full value of AI adoption and integration. To recover the per-use-case effect, we estimate two-stage least squares (2SLS) specifications that instrument firm-level AI usage with randomized treatment assignment.

The first stage estimates the effect of treatment on AI use:

$$AIUseCases_i = \pi_0 + \pi_1 \, Treatment_i + \delta_{s(i)} + \nu_i \tag{3}$$

The second stage uses predicted AI use to estimate its effect on firm outcomes:

$$y_i = \beta_3 \, \widehat{AIUseCases}_i + \delta_{s(i)} + \varepsilon_i \tag{4}$$

Because treatment was randomly assigned, it provides exogenous variation in AI adoption. Assuming the exclusion restriction that treatment affects outcomes only through its effect on AI use, the 2SLS estimate β_3 identifies a local average treatment effect (LATE), corresponding to the

effect of AI use cases for firms whose AI use was induced by the treatment.

5 Results

5.1 Do treated firms change how they use AI?

We begin by examining whether the treatment changes how firms use AI. Each week, ventures reported any new ways they were using AI in their firm, from using Claude Code to build an automated report writer, to deploying Manus AI as a clinical affairs specialist, to prototyping an AI-driven scenario engine in an HR dashboard (see Appendix Table A8 for additional examples). These self-reported use cases, validated by independent human coders, are our primary measure of how broadly firms map AI into their production processes.

As shown in Figure 5, treatment shifts the entire distribution of AI use cases. Panel (a) displays the distribution of cumulative post-baseline use cases: the modal control firm reports zero or one use case, while treated firms are shifted substantially to the right. On average, treated firms reported 8.8 cumulative AI use cases compared to 6.1 for control firms, a difference of 2.7 use cases or roughly a 44% increase relative to the control mean. Panel (b) shows that this gap emerges after treatment begins in week 3 and grows steadily over the program, consistent with cumulative discovery rather than a one-time shift. Our primary specification with strata fixed effects yields an estimate of +2.69 additional AI use cases ($p < 0.01$; Table 3).

Treatment also expands the breadth of AI adoption. We classify reported applications into seven functional domains: technical infrastructure, product development, marketing, research/legal/writing, product/strategy design, business operations, and sales (details in Appendix 7.3). Treated firms adopt AI across 0.84 more distinct functional categories compared to the control ($p < 0.01$; Table 3), suggesting that the intervention does not simply intensify existing uses but expands the range of business functions in which firms apply AI. This breadth matters: when firm production is complementary, applying AI to only one function can leave bottlenecks intact and mute performance effects (Gans and Goldfarb, 2026).

Moreover, treated firms differ significantly in the types of use cases adopted. Figure 6 breaks out average use cases per venture across seven functional categories. The largest and most significant differences are in product development, where treated firms use AI to prototype and build products; product/strategy design, where firms embed AI as a core feature of the product itself (Kim and Koning, 2026) and use AI to inform high-level strategic decisions; and business operations, where

firms automate internal workflows like email triage and process coordination. The differences are relatively smaller in categories like research/legal/writing and sales, where AI use looks more similar across groups. The pattern is telling: treatment shifts AI usage most in areas where founders must rethink how work is organized, not just layer a chatbot onto an existing task. Put differently, the treatment expands both AI as a *process* input—where AI helps the team build faster—and AI as a *product* output—where AI capabilities are delivered directly to customers. That effects concentrate in both suggests the intervention expanded not just how firms build but what they build.

These results reflect differences in AI use cases, not in the AI tools firms use. Appendix Figure A5 displays AI tool adoption across treated and control ventures: all major tools are used broadly by both groups, though treated firms do adopt more tools overall. The treatment gets firms to use AI more, but as the use case results above show—and as we discuss in the sections that follow—it also changes how firms reorganize the rest of their production processes around it. The binding constraint is not access to the technology, both groups use multiple AI tools, but knowing where to deploy it and how to restructure the firm to benefit from them.

5.2 Does treatment change what tasks firms complete?

These changes in AI use lead to meaningful differences in venture activity. Each week, ventures reported the concrete actions they took to advance their startup, from building an MVP to pitching investors to onboarding beta testers (see Appendix 7.2.2 for coding details and examples). Aggregating across the post-baseline period, treated firms completed significantly more tasks than control firms—20.9 tasks on average compared to a control mean of 18.5 (Figure 7). Our primary specification with strata fixed effects yields an estimate of +2.3 tasks ($p < 0.05$; Table 3). The treatment effect grows over time, consistent with cumulative gains from increased AI use (Figure 7).

To understand where these gains occur, we had two independent human coders classify each task as either *internal*—activities performed within the firm like developing a product, building a landing page, or creating financial projections—or *external*—activities involving people outside the firm like customer interviews, investor pitches, or sales outreach (Appendix 7.2.2). A third coder resolved disagreements.

The treatment effect is driven almost entirely by internal tasks. Treated firms complete approximately 2.2 more internal tasks relative to a control mean of 9.8, a 22% increase ($p < 0.01$), with no statistically significant difference in external task completion (Table 3). This pattern makes sense given our treatment design: the accelerator provided identical opportunities to meet investors

and potential customers across all firms, so we would not expect external tasks to differ. What AI changes is how much firms can build—and the internal task results suggest they are building more.

Founders' own reflections point in the same direction. One treated founder reflected: "This mindset shift fundamentally changed how we build at [REDACTED]. I began using AI tools not as a replacement for expertise but as a force multiplier ... a way to perform the 'analyst or intern' version of a task that I can then refine further." Another explained: "In just a few hours I was able to produce what previously cost $1,000 from an outsourced dev team ... We realized we will need even fewer staff than we initially thought, because AI allows us to do in house what we assumed required outsourcing."

5.3 Does treatment impact firm performance?

Measuring firm performance is challenging because early-stage ventures can make meaningful progress without yet generating revenue, and costs are often unclear. In practice, investors assess progress through a set of milestones, and we take a similar approach. We examine whether firms have launched a product, acquired customers, generated revenue, and raised investment—each capturing a different dimension of traction (Table 4).

We find that treatment induces meaningful improvements across these milestones. Overall, treated firms' Venture Progress Index increases by 0.43 standard deviations ($p < 0.05$). Appendix Figure A4 shows the distribution of the underlying index, which sums the four binary milestones (range 0–4). The histogram reveals that treatment both reduces the share of ventures achieving zero milestones and shifts substantially more ventures to achieving all four. This suggests that AI is not simply helping firms produce a prototype that nobody wants; rather, treated firms are progressing along a range of milestones—from product launch through customer acquisition and revenue generation—consistent with genuine advancement. Turning to the individual components, treated firms are 4.7 percentage points (6%) more likely to have launched a product ($p < 0.1$) and 11 percentage points (18%) more likely to have acquired paying customers ($p < 0.01$). Table A2 further supports these results: treated firms are 8.9 percentage points more likely to transition from having no product to launching one, and 14.4 percentage points more likely to transition from having no customers to acquiring their first. Treated firms generate 1.9× higher total revenue on average ($\log(x+1)$ coefficient 0.653, $p < 0.1$); conditional on earning revenue, the effect rises to 2.2× (log coefficient 0.783, $p < 0.05$). Both results are robust to alternative transformations, including the inverse hyperbolic sine and various winsorization thresholds.(Appendix Table A3). We observe lim-

ited average effects on investment outcomes, with no statistically significant impact on the likelihood of raising external capital or on total investment amounts (Table 4; Appendix Table A4).

Notably, these gains do not appear to be driven by increased pivoting or by ventures shutting down and restarting with a new idea. The coefficient on pivoting is positive but statistically insignificant (0.057, $p > 0.1$), and the coefficient on shutdown is slightly negative and insignificant (-0.005, $p > 0.1$; Table A2).

Treatment improves firm performance on average, but for high-growth startups we are often especially interested in how a technology shifts the distribution of outcomes—and in particular, whether it produces right-tail winners (Kerr, Nanda, and Rhodes-Kropf, 2014; Koning, Hasan, and Chatterji, 2022). We estimate quantile regressions that allow us to examine treatment effects at each point in the outcome distribution. As shown in Figure 8, the revenue gains from treatment are small through most of the distribution but rise sharply in the upper tail, with the largest effects emerging at the 90th and 95th percentiles. These quantile results also help explain the noisiness of the average log revenue estimate: the treatment effect is driven by a subset of firms that break out, rather than a uniform shift. Turning to investment, we find a similar pattern: the right tail of the investment distribution is wider for treated firms (Figure 9), with the highest-performing treated firms raising substantially more capital than their control counterparts. Full quantile regression tables are reported in the Appendix (Tables A10–A13); these results are robust to winsorizing at the 99th percentile, confirming they are not driven by a single extreme outlier.

These right-tail gains are consistent with the mapping problem framework introduced above: when founders learn to apply AI across their production processes, it can alleviate binding bottlenecks that unlock disproportionate value (Kremer, 1993; Gans and Goldfarb, 2026). The pattern suggests that AI lifts the ceiling of what the most promising ventures can achieve, rather than primarily lowering entry barriers or modestly improving outcomes for marginal firms.

5.4 Do treated firms think differently about the inputs they need?

A striking pattern in our data concerns not what treated firms produce, but what they consume. Given faster growth and increased traction, one might expect treated firms to demand more external resources. Yet we observe the opposite. Treated firms report just over $220,000 less in capital demand relative to control firms, a 39.5% decrease ($p < 0.05$), with no corresponding increase in labor demand (Table 5). This capital demand result is robust to alternative transformations and winsorization levels (Appendix Table A5).

These averages are not to say that AI-using firms do not need to raise capital. When we examine the quantile distribution of investment raised, we find that treated firms in the upper tail raise substantially more funding than their control counterparts (Figure 9). However, this conflates two things: the firm's need for capital and its ability to attract it, since faster-progressing ventures may simply be more appealing to investors. Consistent with the idea that AI lowers firms' perceived need for capital, when we turn to the quantile distribution of capital demand (Figure 10), we see that the largest reductions in capital ask come from the right tail, above the 60th percentile. Taken together, our results suggest that treated firms think they can do more with less, consistent with the idea that AI may fundamentally change what firms can produce with a given set of inputs. Full quantile regression tables for capital demand are reported in the Appendix (Tables A14–A15); these patterns are robust to winsorizing at the 99th percentile.

5.5 Alternative explanations for the performance effects

A natural concern is that treatment effects reflect heightened motivation or effort rather than the specific content of the intervention—a Hawthorne-style effect. We think this is unlikely. If control firms were simply less motivated, they should be more likely to attrit. But as shown in Appendix Table A1, weekly survey response rates are nearly identical across groups throughout the program, with no statistically significant differences at any point; if anything, control firms were slightly more likely to complete all surveys without missing a week (81.2% vs. 80.8%). We also find no differences in sentiment about the program: Appendix Figure A3 reports sentiment analysis of founders' open-ended text responses using both VADER and RoBERTa classifiers, with no meaningful differences in affect between treated and control firms.

A second concern is spillovers between treatment and control firms. We took several steps to prevent this. Treatment and control firms were assigned to fully separate groups on the program's online platform with no ability to view each other's content, and live chat was disabled during joint sessions with outside speakers. When asked at endline about sessions they wished they could have attended, zero control founders mentioned treatment-specific sessions. Attendance records confirm that no control firms attended treatment-only workshops. Additional details are provided in Appendix 7.1.

5.6 How do treated firms reorganize around AI?

Having ruled out these alternative explanations, what evidence do we have that firms actually reorganized around AI? The results above show that treated firms use more AI, complete more tasks, perform better, and demand less capital. But what does this reorganization actually look like? Appendix Table A9 presents illustrative examples drawn from ventures' weekly progress reports, in which they described how they were using AI and what they accomplished each week. Several patterns emerge that connect directly to the mechanisms outlined in our treatment vignettes (Figure 1–Figure 4).

First, treated firms use AI to rapidly prototype and build products that would otherwise require specialized talent. A tender-matching platform prototyped an end-to-end pipeline from tender classification through compliance checking and bid pricing, going from zero to $40K in revenue and four paying customers during the program—without hiring the technical talent that would conventionally be needed to build this pipeline.

Second, we see firms reorganize internal workflows by replacing sequential, labor-intensive processes with AI. A field services venture rebuilt its full operations chain—from an intent classifier replacing a dispatcher to a payment-reminder engine replacing collections staff—compressing five roles into AI modules that self-improve from operator feedback. A fintech with 2,500 existing customers moved KYC screening, its call center, and even hiring interviews to AI, growing its customer base by 20% while reducing costs.

Third, these changes enable firms to operate at scale or in markets that were previously uneconomical. A maternal health venture replaced hour-long clinical intake sessions with a WhatsApp chatbot that conducts intake and classifies risk in under eight minutes, making it viable to screen patients in low-resource settings where trained health workers are scarce. A trade quality-control startup built a multi-modal pipeline—image forensics, compliance report drafting, location verification—that reduced manual inspection by 70% and grew the venture from two to forty customers during the program. A medical transport company shifted from reactive human dispatch to AI-driven routing by patient urgency and traffic, with predictive positioning of drivers before hospital discharge peaks, reaching 160 rides per week with 60% of bookings handled entirely by AI. In each case, the gains came not from automating a single task but from reorganizing multiple complementary activities so that improvements in one step were not bottlenecked by constraints in the next.

5.7 Who benefits from reorganizing around AI?

A natural question is whether the benefits of reorganizing around AI accrue broadly or only to certain types of founders and firms. We examine heterogeneity along two dimensions: baseline venture performance and founder technical background.

We first examine heterogeneity by baseline firm performance (Appendix Table A7). On our primary outcome, the Venture Progress Index, the interaction between treatment and high traction is small and insignificant, indicating no differential effects by baseline performance. The customer acquisition effect is somewhat concentrated among firms that had not yet reached that milestone (-17.8 pp interaction, $p < 0.05$), but otherwise the treatment helps across the traction distribution.

We next ask whether technically trained founders benefit more. If the intervention primarily taught technical skills—how to prompt, how to code with AI—we would expect founders with technical backgrounds to benefit less, since they likely already possess these capabilities. Instead, we find no significant differential effects by technical background: the interaction between treatment and having a technical founder is small and statistically insignificant across all outcomes (Appendix Table A6).

5.8 Estimating the firm-level returns to AI use

Our ITT estimates capture the effect of helping firms solve the mapping problem, but the intervention also lets us estimate the returns to AI use itself. Because our treatment shifts the number of AI use cases firms adopt, we can instrument firm-level AI usage with randomized treatment assignment to recover the marginal value of each additional use case (Table 6). This approach identifies the local average treatment effect of an additional AI use case for firms whose adoption was shifted by the treatment—providing an estimate of AI's returns that extends beyond the mapping friction per se. It requires the assumption that treatment affects outcomes only through its effect on AI use, which we discuss below.

The first stage is strong, with F-statistics exceeding 33, confirming that treatment substantially shifts AI use. The second-stage estimates suggest that each additional AI use case increases the number of tasks completed by 0.85 ($p < 0.01$), raises the Venture Progress Index by 0.15 standard deviations ($p < 0.05$), increases the probability of having customers by 3.9 percentage points ($p < 0.01$), increases log revenue by 0.23 ($p < 0.1$), and reduces log capital demand by 0.17 ($p < 0.1$). These estimates are consistent with the ITT results: the treatment-induced 2.7 additional use cases,

valued at 0.15 SD each, predict a 0.40 SD increase in the Venture Progress Index—close to the ITT estimate of 0.43 SD.

To put the magnitude in perspective, control firms already average 6.1 AI use cases and treated firms average 8.8. If we take the per-use-case LATE at face value and extrapolate linearly, a firm moving from zero AI use cases to the treatment group mean of 8.8 would see a 1.3 SD increase in the Venture Progress Index and a roughly 34 percentage point increase in the probability of acquiring paying customers. These are obviously relatively heroic extrapolations—the LATE is identified locally around the treatment-induced margin. But they provide a useful benchmark: the cumulative value of mapping AI broadly across a firm's production process appears to be substantial, and the treatment-induced margin of 2.7 additional use cases captures only a fraction of the value of AI.

6 Discussion

We study 515 high-growth startups in a field experiment that encouraged treated firms to explore how AI could reshape their production processes. Treated firms identified more AI use cases, completed more tasks, and were more likely to acquire paying customers and grow revenue—with the largest gains in the upper tail, consistent with with the idea that reorganizing production around AI can expand the upper range of what ventures can achieve. Treated firms also demanded hundreds of thousands of dollars less in external capital, with no change in workforce size or labor demand, consistent with a view that AI enables firms to produce more with fewer resources. These effects are broad-based across technical backgrounds and performance levels, suggesting that the mapping problem is a key constraint on firm-level value creation.

Taken together, these results offer a partial resolution to the disconnect between AI's demonstrated capabilities and its limited appearance in firm-level and aggregate productivity statistics. Our results suggest that the bottleneck is not the technology — it is the managerial challenge of discovering where the technology creates value within a firm's production process. This friction likely extends beyond AI. Any general-purpose technology whose applications are vast, uneven, and hard to predict ex ante introduces a similar mapping challenge of technology adoption under unknown complements.

Furthermore, we observe a wide variety of patterns across firms in how they integrate AI, rather than a single best practice. This is consistent with the idea that the ways in which AI maps into

production may be firm-specific. If this heterogeneity is present even among early-stage ventures with minimal organizational complexity, the challenge is likely greater still for established firms.

Our findings carry a practical implication for how policymakers and firms think about AI adoption. Much of the current conversation focuses on access: subsidizing tools or compute, distributing API credits, or training workers to use AI. Our results suggest that access alone is not enough. Control firms in our experiment had the same tools, the same credits, and the same technical training as treated firms—yet they realized substantially less value from AI. What treated firms received instead was information about how to explore reorganizing their production around AI: where to deploy it, how complementary activities must change, and what reorganized production actually looks like in practice. Teaching managers and entrepreneurs how to solve the mapping problem may be at least as important as ensuring they have access to the technology.

Our study has important limitations. The ten-week experimental window, while longer than the majority of experiments testing the impact of generative AI on business outcomes, still captures short-run effects. Whether these performance gains compound or dissipate as ventures mature remains an open question. We study early-stage firms where organizational inertia is low; how the mapping problem evolves as firms grow—as hierarchies deepen, processes calcify, and decision authority becomes distributed—is an important question we cannot address here. Finally, as with all studies of a rapidly evolving technology, the specific ways AI maps into a firm's activities will shift as capabilities expand.

That said, our core finding centers not on a technological friction but a managerial one: the difficulty of discovering where and how to deploy AI within a firm's production process. This friction may, in fact, become more important as AI improves. As models grow more capable, the set of activities they can perform expands, making the search space larger and the mapping problem harder. A firm that has figured out where to use today's AI will likely face the same discovery problem again when next year's models can do more.

More broadly, our findings point to the importance of studying how AI is integrated within firms, not only how it performs on specific tasks. The mapping problem reflects a deeper information-processing constraint: firms persistently struggle to search for the information needed to realize value from new technologies. Our results suggest that AI does not eliminate these constraints but creates a new and consequential domain in which they bind. As AI capabilities expand, the central question for firms may be less about the technology itself and more about how they search for where and how to integrate it—and how to reorganize production once they do.

References

Agrawal, Ajay, Arnaldo Camuffo, Alfonso Gambardella, Joshua Gans, Erin L. Scott, and Scott Stern, editors. 2026. *Bayesian Entrepreneurship*. Cambridge, MA: The MIT Press.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial intelligence, firm growth, and product innovation." *Journal of Financial Economics* 151:103745. URL `https://www.sciencedirect.com/science/article/pii/S0304405X2300185X`.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does management matter? Evidence from India." *The Quarterly journal of economics* 128 (1):1–51.

Bresnahan, Timothy, Shane Greenstein, David Brownstone, and Ken Flamm. 1996. "Technical progress and co-invention in computing and in the uses of computers." *Brookings Papers on Economic Activity. Microeconomics* 1996:1–83.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond. 2023. "Generative AI at work." NBER Working Paper 31161.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2018. "The Productivity J-Curve: How Intangibles Complement General Purpose Technologies." NBER Working Paper 25148, National Bureau of Economic Research.

———. 2021. "The productivity J-curve: How intangibles complement general purpose technologies." *American Economic Journal: Macroeconomics* 13 (1):333–372.

Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella, and Chiara Spina. 2020. "A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial." *Management Science* 66 (2):564–586.

Chatterji, Aaron, Solène Delecourt, Sharique Hasan, and Rembrand Koning. 2019. "When does advice impact startup performance?" *Strategic Management Journal* 40 (3):331–356.

Cohen, Susan, Daniel C Fehder, Yael V Hochberg, and Fiona Murray. 2019. "The design of startup accelerators." *Research Policy* 48 (7):1781–1797.

Csaszar, Felipe A, Harsh Ketkar, and Hyunjin Kim. 2024. "Artificial intelligence and strategic decision-making: Evidence from entrepreneurs and investors." *Strategy Science* 9 (4):322–345.

David, Paul A. 1990. "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox." *American Economic Review* 80 (2):355–361.

Dell'Acqua, Fabrizio, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2023. "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality." SSRN 4573321.

Demirer, Mert, J Horton, Nicole Immorlica, Brendan Lucier, and Peyman Shahidi. 2025. "Chaining Tasks, Redefining Work: A Theory of AI Automation." Tech. rep., Working Paper.

Dimitriadis, Stefan and Rembrand Koning. 2022. "Social skills improve business performance: Evidence from a randomized control trial with entrepreneurs in Togo." *Management Science* 68 (12):8635–8657.

Dushnitsky, Gary and Bryan K. Stroube. 2021. "Low-code entrepreneurship: Shopify and the alternative path to growth." *Journal of Business Venturing Insights* 16:e00251.

Ewens, Michael, Ramana Nanda, and Matthew Rhodes-Kropf. 2018. "Cost of experimentation and the evolution of venture capital." *Journal of Financial Economics* 128 (3):422–442.

Fleming, Lee. 2001. "Recombinant uncertainty in technological search." *Management science* 47 (1):117–132.

Gans, Joshua S and Avi Goldfarb. 2026. "O-Ring Automation." Tech. rep., National Bureau of Economic Research.

Gavetti, Giovanni and Daniel A. Levinthal. 2000. "Looking Forward and Looking Backward: Cognitive and Experiential Search." *Administrative Science Quarterly* 45 (1):113–137.

Gavetti, Giovanni, Daniel A. Levinthal, and Jan W. Rivkin. 2005. "Strategy Making in Novel and Complex Worlds: The Power of Analogy." *Strategic Management Journal* 26 (8):691–712.

Hendrycks, Dan, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. "Measuring Massive Multitask Language Understanding." In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Kerr, William R, Ramana Nanda, and Matthew Rhodes-Kropf. 2014. "Entrepreneurship as experimentation." *Journal of Economic Perspectives* 28 (3):25–48.

Kim, Hyunjin. 2024. "The value of competitor information: Evidence from a field experiment." Management Science (forthcoming).

Kim, Hyunjin, Edward L Glaeser, Andrew Hillis, Scott Duke Kominers, and Michael Luca. 2024. "Decision authority and the returns to algorithms." *Strategic Management Journal* 45 (4):619–648.

Kim, Hyunjin and Rembrand Koning. 2026. "AI-Native FIrms." *Working Paper* .

Kim, Hyunjin and Nety Wu. 2025. "Discovering Alternative Strategies: Experimental Evidence on the Impact of Frameworks." Working Paper.

Koning, Rem, Nicole Tempest Keller, and Hyunjin Kim. 2025. "Gamma: Slides in the Blink of AI." Case study, Harvard Business School.

Koning, Rem, Hyunjin Kim, and Nicole Tempest Keller. 2025. "Fazeshift: AI for AR." Case study, Harvard Business School.

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

Kremer, Michael. 1993. "The O-ring theory of economic development." *The quarterly journal of economics* 108 (3):551–575.

Lazear, Edward P. 2004. "Balanced skills and entrepreneurship." *American Economic Review* 94 (2):208–211.

Levinthal, Daniel A. and James G. March. 1993. "The Myopia of Learning." *Strategic Management Journal* 14 (S2):95–112.

McElheran, Kristina, J Frank Li, Erik Brynjolfsson, Zachary Kroff, Emin Dinlersoz, Lucia S Foster, and Nikolas Zolas. 2023. "AI Adoption in America: Who, What, and Where." Tech. rep., National Bureau of Economic Research.

Milgrom, Paul and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80 (3):511–528.

Miserendino, Samuel, Michael Wang, Tejal Patwardhan, and Johannes Heidecke. 2025. "SWE-Lancer: Can Frontier LLMs Earn $1 Million from Real-World Freelance Software Engineering?" In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, *PMLR*, vol. 267.

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

Otis, Nicholas, Rowan Clarke, Solene Delecourt, David Holtz, and Rembrand Koning. 2024. "The uneven impact of generative AI on entrepreneurial performance." .

Patwardhan, Tejal, Rachel Dias, Erica Proehl, Grace Kim, Michael Wang, Olivia Watkins, Samuel P. Fishman, Muhannad Aljubeh, Pranav Thacker et al. 2025. "GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks." Working paper, OpenAI.

Phan, Long, Alice Gatti, Ziwen Han, Nathaniel Li et al. 2026. "Humanity's Last Exam." Working paper, Center for AI Safety / Scale AI.

Reimers, Imke and Joel Waldfogel. 2026. "AI and the Quantity and Quality of Creative Products: Have LLMs Boosted Creation of Valuable Books?" Tech. rep., National Bureau of Economic Research.

Seamans, Robert and Manav Raj. 2018. "AI, Labor, Productivity and the Need for Firm-Level Data." Working Paper 24239, National Bureau of Economic Research. URL `https://www.nber.org/papers/w24239`.

Siggelkow, Nicolaj. 2002. "Misperceiving Interactions Among Complements and Substitutes: Organizational Consequences." *Management Science* 48 (7):900–916.

Starace, Giulio, Oliver Jaffe, Dane Sherburn, James Aung, Jun Shern Chan, Leon Maksin, Rachel Dias, Evan Mays, Benjamin Kinsella, Wyatt Thompson, Johannes Heidecke, Amelia Glaese, and Tejal Patwardhan. 2025. "PaperBench: Evaluating AI's Ability to Replicate AI Research." Working paper. ArXiv:2504.01848.

Stuart, Toby E and Joel M Podolny. 1996. "Local search and the evolution of technological capabilities." *Strategic management journal* 17 (S1):21–38.

Vafa, Keyon, Ashesh Rambachan, and Sendhil Mullainathan. 2024. "Do large language models perform the way people expect? measuring the human generalization function." *arXiv preprint arXiv:2406.01382* .

Tables and Figures

Table 1: AI Founder Sprint: Accelerator Program Overview

Week	Dates	Sessions & Events	Data Collection
Applications	Jul 2025	Online Application	Application Data
Week 1	Aug 25	Sprint Orientation	WPR #1
Week 2	Sep 1–5	Workshop; Technical Training	WPR #2
		Treatment begins	
Week 3	Sep 8–12	***Workshops***; Partner Resources; Technical Training	WPR #3
Week 4	Sep 15–19	***Workshops***; Partner Resources; Technical Training	WPR #4
Week 5	Sep 22–26	***Workshops***; Partner Resources; Technical Training	WPR #5
Week 6	Sep 29–Oct 5	***Workshops***; Technical Training	WPR #6
Week 7	Oct 6–10	VC Week	WPR #7
Week 8	Oct 13–14	***Workshops***	—
Weeks 3–10	Sep 8–Oct 26	***Peer Learning & Office Hours***	—
Week 10	Oct 26	Final Submissions for Demo Day Evaluation	WPR #8
Week 15	Dec 5 & 9	Demo Days; $100K Non-Dilutive Seed Funding	—

- **Weekly Progress Reports (WPR):** Structured reports asking founders to reflect on their work over the last week and report their progress and key metrics.
- **Partner Resources:** ∼$25,000 per venture in API credits, partner tools, and interactive sessions from Google Cloud, OpenAI, NVIDIA, and Manus AI.
- **Technical Training:** Weekly 3-hour hands-on sessions led by the MIT/Harvard Sundai AI Club, covering rapid prototyping, memory, context engineering, RAG, MCPs, agents, agentic workflows, vibe-coding, and engineering teams.
- **VC Week:** Selected founders pitch to leading VCs from across the globe.
- **Demo Days:** Singapore and Abu Dhabi; $100K non-dilutive seed funding.
- ***Workshops:*** 60–90-minute sessions led by Sprint instructors over Zoom featuring case-based discussions and guest speakers. Both groups were invited to workshops and attended and interacted with guest speakers. Treatment workshops included additional information exploring how to reorganize the firm around AI, delivered through case studies, slides, and fireside chats.
- ***Peer Learning & Office Hours:*** Required weekly peer group meetings (4–8 founders with a venture guide) plus opt-in office hours with mentors and Sprint instructors. These sessions enabled founders to discuss and work together to implement the workshops' content and so vary across treatment and control.

Notes: ***Red bold italic text*** indicates program components whose content varied between experimental conditions.

Table 2: Baseline Balance Table

Variable	Control	Treatment	SD	Min	P25	Median	P75	Max	N	Diff.	p-value
DEMOGRAPHICS											
Male	0.80	0.74	0.44	0.00	1.00	1.00	1.00	2.00	515	-0.06	0.1301
Education Level	2.69	2.65	0.63	1.00	2.00	3.00	3.00	4.00	514	-0.04	0.4310
Work Experience (years)	12.87	13.04	10.40	0.00	6.75	10.50	18.00	154.00	504	+0.16	0.8592
Full-Time on Venture	0.84	0.88	0.35	0.00	1.00	1.00	1.00	1.00	515	+0.04	0.1937
TEAM CHARACTERISTICS											
Team Size	5.24	4.95	5.09	1.00	3.00	4.00	6.00	48.00	515	-0.29	0.5191
Year Founded	2024.14	2024.05	1.15	2015.00	2023.50	2024.00	2025.00	2025.00	515	-0.08	0.4124
Full-Time Employees	3.50	3.19	4.66	0.00	1.00	2.00	4.00	45.00	488	-0.31	0.4621
BASELINE TRACTION – BINARY											
Product Launched	0.52	0.53	0.50	0.00	0.00	1.00	1.00	1.00	515	+0.01	0.8858
Has Customers	0.57	0.54	0.50	0.00	0.00	1.00	1.00	1.00	515	-0.02	0.5815
Has Revenue	0.33	0.36	0.48	0.00	0.00	0.00	1.00	1.00	515	+0.04	0.3685
Has Investment	0.34	0.37	0.48	0.00	0.00	0.00	1.00	1.00	515	+0.03	0.4750
BASELINE TRACTION – CONTINUOUS											
Monthly Revenue ($)	2,103.88	3,005.96	10,787.67	0.00	0.00	0.00	200.00	115,000.00	515	+902.09	0.3432
Investment Amount ($)	74,508.47	134,291.31	736,253.57	0.00	0.00	0.00	100.00	15,000,000.00	515	+59,782.84	0.3613
Venture Progress Index (0–4)	1.75	1.80	1.40	0.00	0.00	2.00	3.00	4.00	515	+0.05	0.6852
BASELINE AI USAGE											
AI Use Cases	3.23	3.40	2.22	0.00	1.00	3.00	5.00	12.00	515	+0.17	0.4000
Number of AI Tools	5.10	4.90	3.24	0.00	2.00	5.00	7.00	13.00	515	-0.19	0.4971
Joint F-test[a]									515	0.982	0.4649
Joint F-test[b]									478	0.607	0.8699
Joint F-test[c]									507	0.926	0.5208

Notes: All baseline variables were collected during the application phase, prior to randomization. To ensure that participants were at an appropriate stage to benefit from the intervention, the study primarily targeted early-stage ventures. Only 13 ventures in our sample were founded before 2022. A small number of ventures entered through referrals or special cases where the screening criteria did not apply, resulting in a minimum founding year of 2015. Education is coded as 1 = high school, 2 = bachelor's degree, 3 = master's or MBA, and 4 = PhD; a small number of free-text responses (e.g. diplomas, certificates, professional or postgraduate degrees) were recoded into the closest category. The Venture Progress Index is a composite measure aggregating four binary indicators: Product Launched, Has Customers, Has Revenue, and Has Investment.

Joint F-tests: [a]Full sample, complete-data variables only ($n = 515$). [b]Full sample, all variables with listwise deletion ($n = 478$). [c]Post-baseline sample (at least 1 post-baseline observation), complete-data variables only ($n = 507$).

Table 3: Treatment Effects on AI Usage and Tasks Done

Panel A: AI Usage		
	Total AI Use Cases	Breadth: Distinct Categories
	(1)	(2)
Treatment	2.686***	0.839***
	(0.492)	(0.162)
Strata FE	Yes	Yes
Observations	507	507
Control Mean	6.054	2.926
R^2	0.147	0.144

Panel B: Tasks Done			
	Number of Tasks Done	Number of External Tasks	Number of Internal Tasks
	(3)	(4)	(5)
Treatment	2.276**	0.076	2.201***
	(1.030)	(0.636)	(0.667)
Strata FE	Yes	Yes	Yes
Observations	507	507	507
Control Mean	18.518	8.739	9.778
R^2	0.101	0.085	0.100

Notes: Each column reports the treatment effect from an OLS regression with strata fixed effects and robust standard errors at the firm level. The sample includes the 507 ventures that submitted at least one post-baseline weekly progress report. Panel A reports treatment effects on AI adoption outcomes. Column (1) reports the effect on total cumulative AI use cases post-baseline. Column (2) reports the effect on breadth, measured as the number of distinct use case categories (out of seven) in which a venture adopted at least one AI application. Panel B reports treatment effects on task completion outcomes. Column (3) reports the effect on total tasks completed post-baseline. Columns (4) and (5) decompose this outcome into external tasks – tasks involving people or organizations outside the venture (e.g. customer interviews, investor meetings, sales outreach) – and internal tasks, which are performed within the venture without external interaction (e.g. building, prototyping, research). Standard errors are reported in parentheses.
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 4: Treatment Effects on Firm Performance

	Venture Progress Index (1)	Product Launched (2)	Has Customers (3)	Has Revenue (4)	Total Revenue (Log) (5)	Has Investment (6)	Total Investment (Log) (7)
Treatment	0.429**	0.047*	0.110***	-0.027	0.653*	0.039	0.436
	(0.203)	(0.025)	(0.041)	(0.039)	(0.382)	(0.038)	(0.458)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	479	479	479	479	479	479	479
Control Mean	-0.263	0.885	0.626	0.481	4.171	0.296	4.152
R^2	0.372	0.130	0.174	0.342	0.377	0.284	0.280

Notes: Standard errors are reported in parentheses. All outcomes are measured at Week 10. The Venture Progress Index is a standardized composite measure aggregating four binary indicators: Product Launched, Has Customers, Has Revenue, and Has Investment. Total Revenue and Total Investment are cumulative self-reported measures and are transformed using a $\log(x+1)$ transformation. Results are broadly consistent across alternative transformations of these outcomes; see Tables A3 and A4 for details. All specifications include strata fixed effects.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Treatment Effects on Capital and Labor Demand

	Capital Ask (Level, $1,000s) (1)	Capital Ask (Log) (2)	Labor Demand (3)	Tech-Related Job Postings (4)
Treatment	-223.955**	-0.502**	-0.071	-0.047
	(108.052)	(0.254)	(0.171)	(0.118)
Strata FE	Yes	Yes	Yes	Yes
Observations	479	479	479	479
Control Mean	850.233	12.514	2.704	1.695
R^2	0.094	0.100	0.054	0.051

Notes: Standard errors are reported in parentheses. Capital Ask (Level, $1,000s) measures the amount of capital the venture would seek to raise from investors on Demo Day, as reported by founders in the Week 10 survey; values are reported in thousands of dollars. Capital Ask (Log) applies the Log(x+1) transformation to the same variable. Results are robust to alternative transformations; see Table A5 for details. Labor Demand measures the total number of job postings reported by the venture at Week 10, where founders were asked to list each role and position they wanted to hire for on Demo Day; each listed position is counted as a distinct posting. Job postings were categorized using keywords into tech roles (engineering, AI specialist, product, design) and non-tech roles (marketing, sales, operations, finance, HR, customer success); Tech-Related Job Postings counts postings in tech categories. All specifications include strata fixed effects.

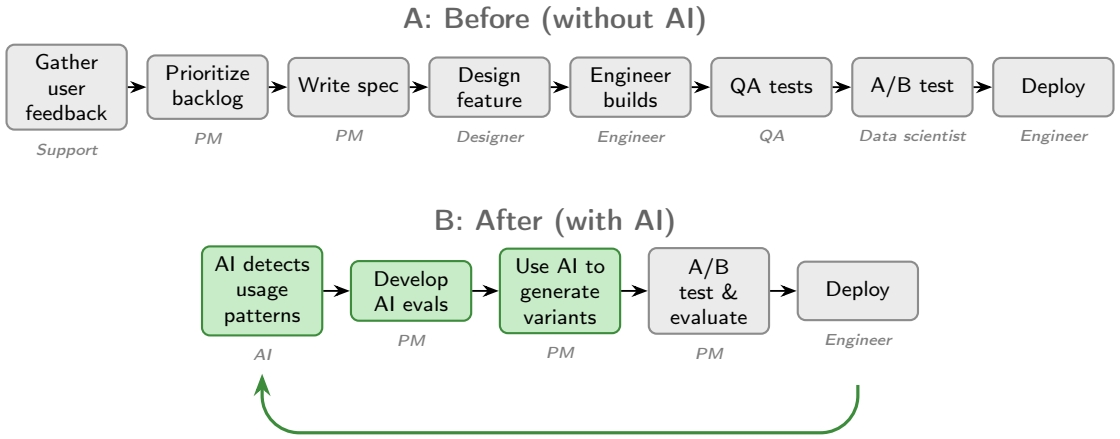
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Effects of AI Use Cases on Key Outcomes (2SLS)

	Number of Tasks Done (1)	Venture Progress Index (2)	Product Launched (3)	Has Customers (4)	Has Revenue (5)	Total Revenue (Log) (6)	Capital Demand (Log) (7)
AI Use Cases	0.847***	0.152**	0.017*	0.039***	-0.010	0.231*	-0.178*
	(0.308)	(0.072)	(0.009)	(0.015)	(0.013)	(0.134)	(0.097)
Strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	507	479	479	479	479	479	479
First-stage F	31.87	33.84	33.84	33.84	33.84	33.84	33.84

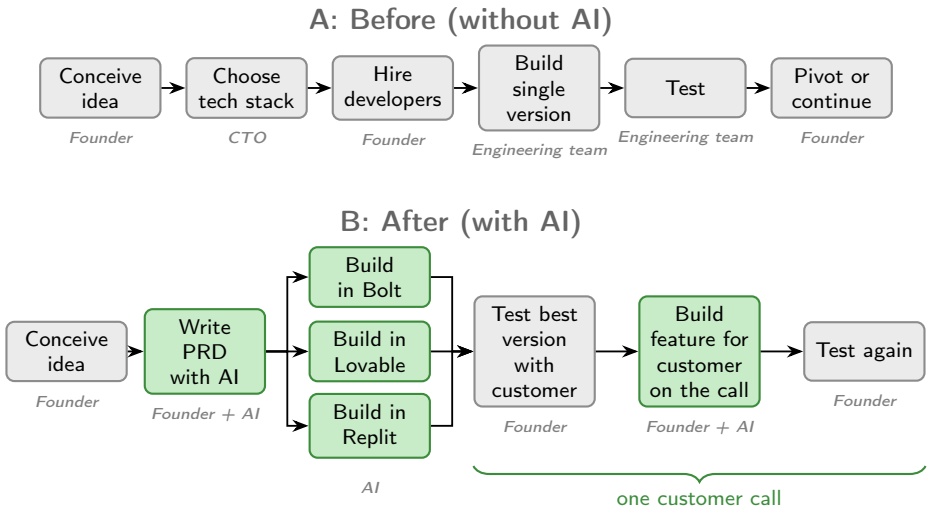
Notes: Standard errors in parentheses. Each column reports a separate 2SLS regression of y_i on instrumented post-baseline AI use cases, using treatment assignment as the instrument. All specifications include strata fixed effects. Outcomes are measured at Week 10, except Number of Tasks Done, which is cumulative across weeks post-baseline. Log outcomes are defined as $\log(x + 1)$. First-stage F-statistics are reported.

Figure 1: Diagram illustrating the Gamma case on AI-native product development.



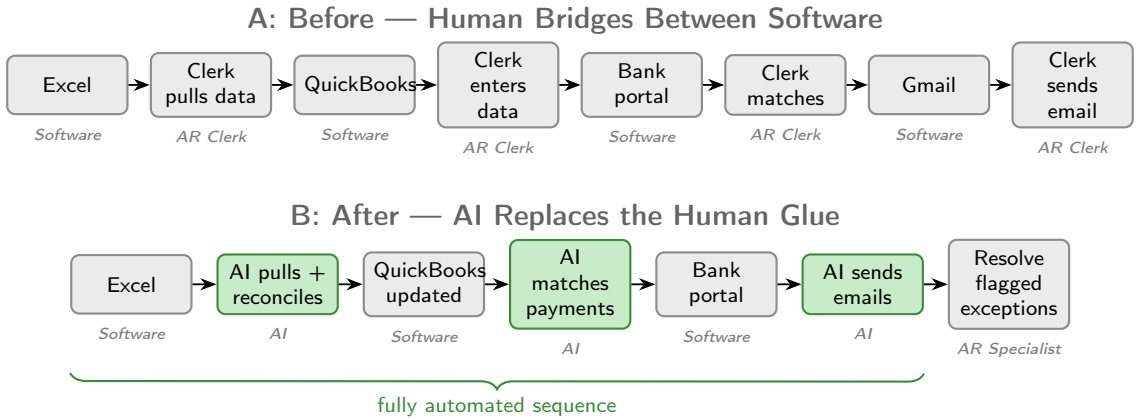
Panel A shows conventional product development: a chain of specialists each handling one step in quarterly batch cycles. Panel B shows how Gamma uses AI to detect usage patterns and generate product variants directly, enabling a single PM to ship features continuously, along with the introduction of a new task, developing AI evals. The difference illustrates how the gains come not from inserting AI into a single step but from rethinking the entire product development process around the technology.

Figure 2: Diagram illustrating the Ryz Labs session on parallel prototyping with AI.



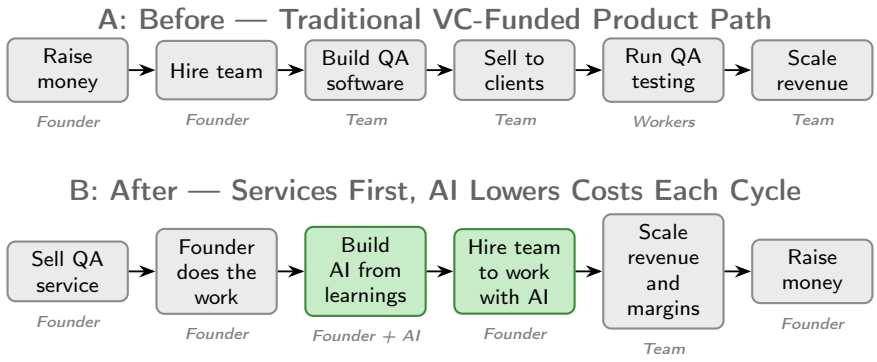
Panel A shows conventional prototyping: a sequential bet on one tech stack, one engineering team, and one build over months. Panel B shows how near-zero build costs allow a founder to launch the same idea across three AI coding tools simultaneously, test the best version with a customer, and iterate live on the call (marked by the brace). AI both overcomes the bottleneck around building variants and can improve testing, unlocking a potential complementarity.

Figure 3: Diagram illustrating the FazeShift case on replacing human glue with AI.



Panel A shows a conventional accounts receivable process: eight steps alternating between software systems and clerks who manually bridge between them. Panel B shows how AI replaces this bridging into a fully automated end-to-end sequence (green bracket), with human intervention only for exceptions. Partial automation preserves the bottleneck—the gains require automating the full sequence, and the remaining role shifts from routine data entry to expert judgment on edge cases.

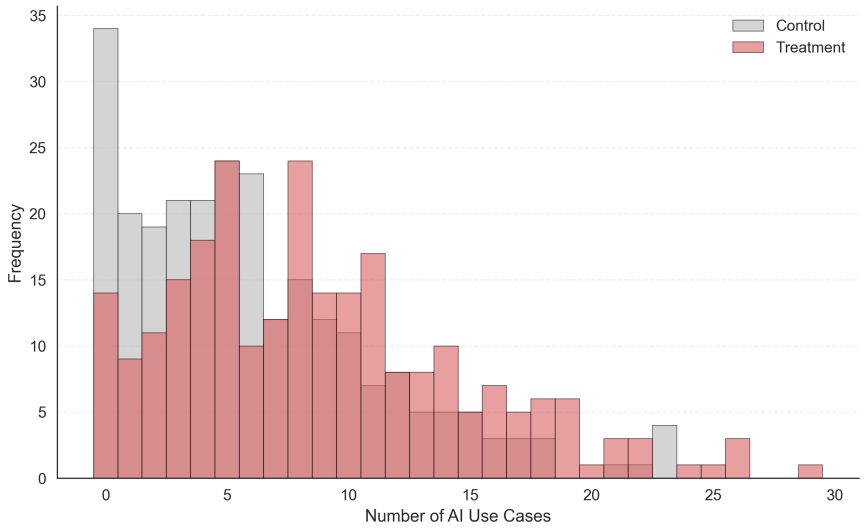
Figure 4: Diagram illustrating the Ranger case on service-to-product business model.



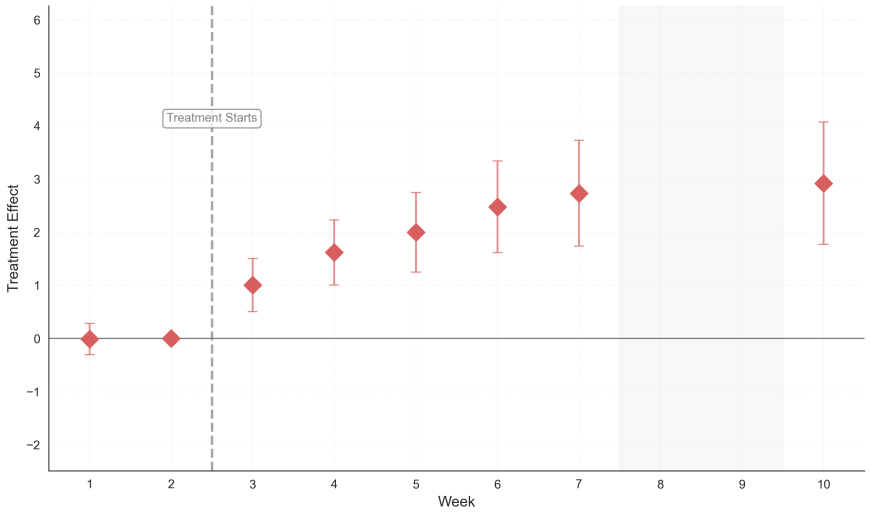
Panel A shows the conventional VC-funded path: all capital is front-loaded before any revenue. Panel B shows how AI enables a services-first alternative where the founder sells from day one, learns which steps are routine, and builds AI to automate delivery. Fundraising moves from the first step to the last, giving optionality on whether and when to raise outside capital.

Figure 5: AI Use Cases

(a) Distribution of Use Cases

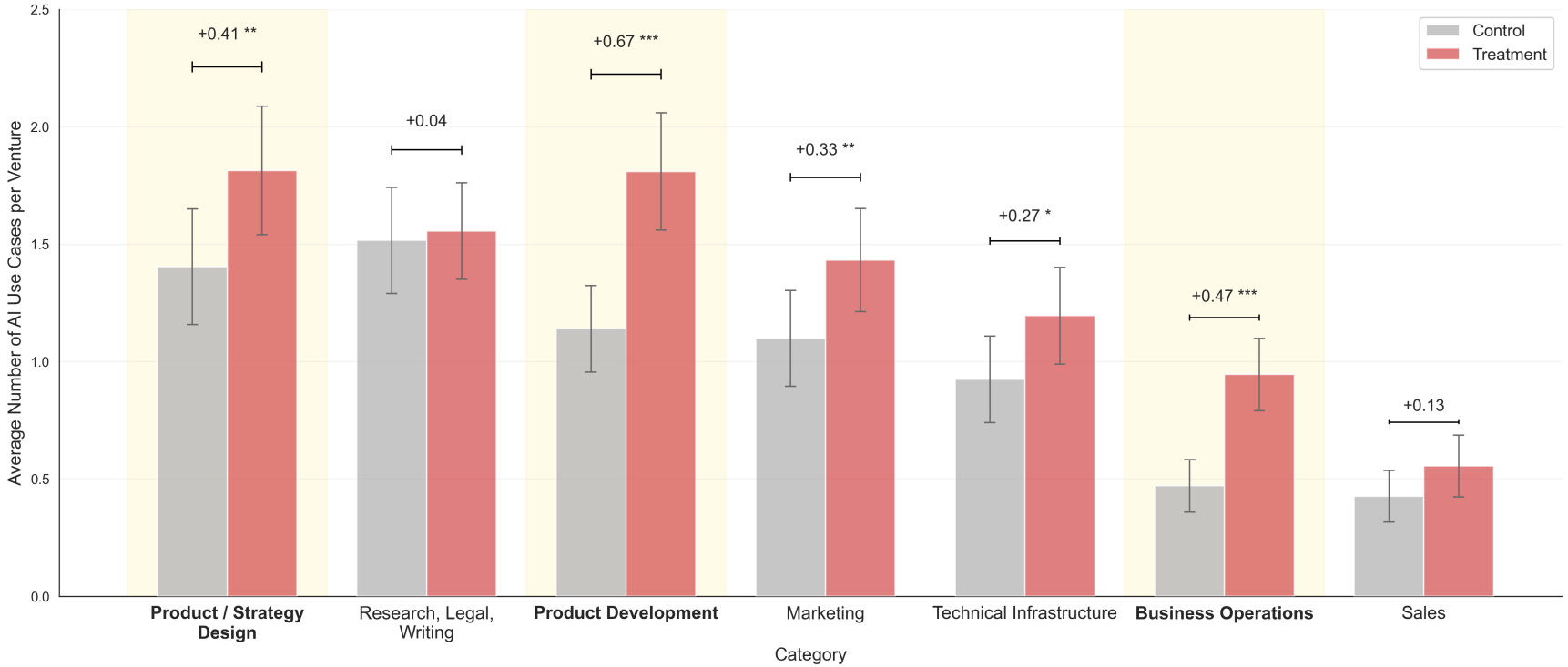


(b) Use Cases Over Time



Notes: Panel (a) displays frequency histograms of cumulative post-baseline AI use cases (weeks 3–10) for treatment and control ventures. Panel (b) reports event study estimates of the treatment effect on cumulative AI use cases by week, from a panel regression of the outcome on treatment × week interactions, week fixed effects, baseline AI use cases, and strata fixed effects, with standard errors clustered at the firm level. The dashed vertical line marks the start of treatment (after week 2). The shaded region indicates weeks 8–9, during which no weekly surveys were administered; week 10 corresponds to the endline survey. Vertical bars denote 95% confidence intervals. $N = 507$ ventures for distribution of use cases; $N = 515$ ventures for use cases over time.

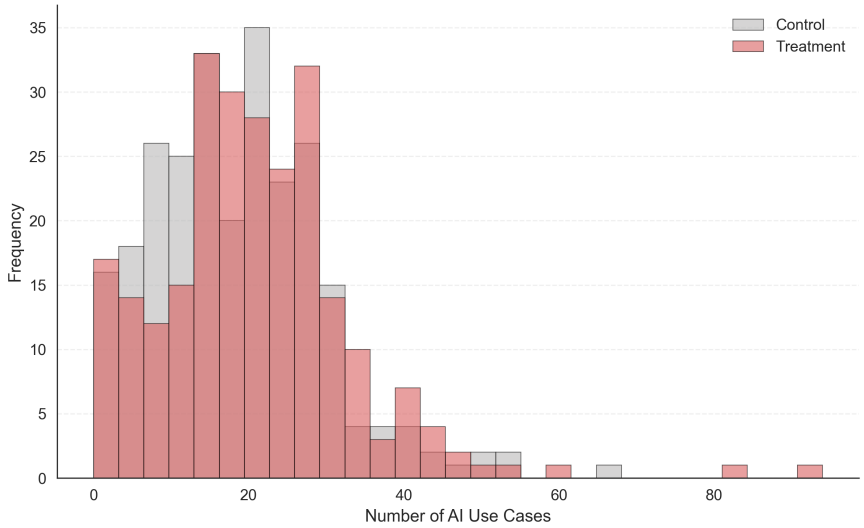
Figure 6: AI Use Case Categories at the Venture Level



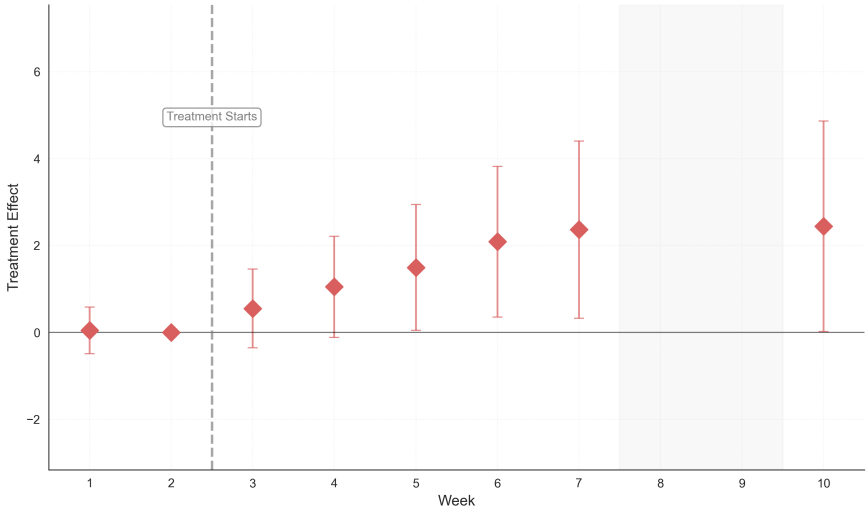
Notes: This figure displays the average number of AI use cases per venture across use case categories. Categories are mutually exclusive and are assigned using an LLM-based classification of submitted use cases with iterative human review. The sample includes only ventures that submitted at least one valid AI use case during the post-baseline period. Founders reported newly implemented AI use cases each week, so counts reflect cumulative use over the reporting period.

Figure 7: Tasks that Ventures Get Done

(a) Distribution of Tasks

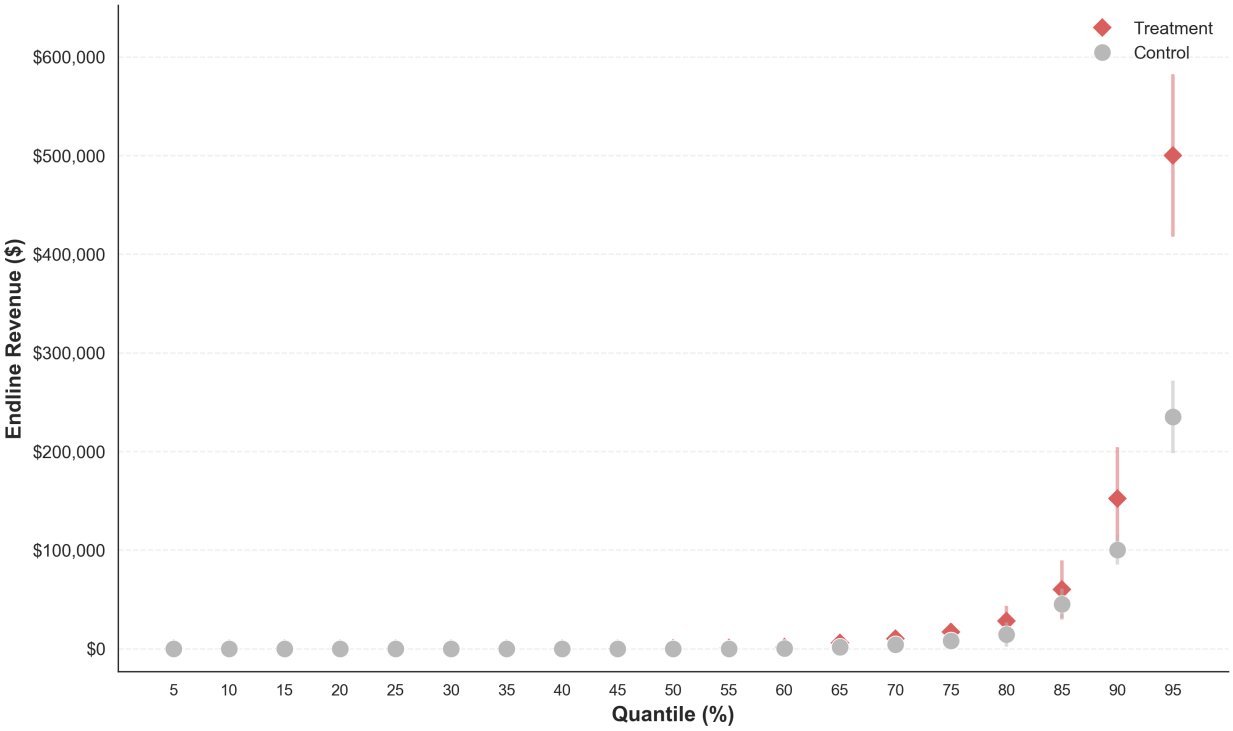


(b) Tasks Over Time



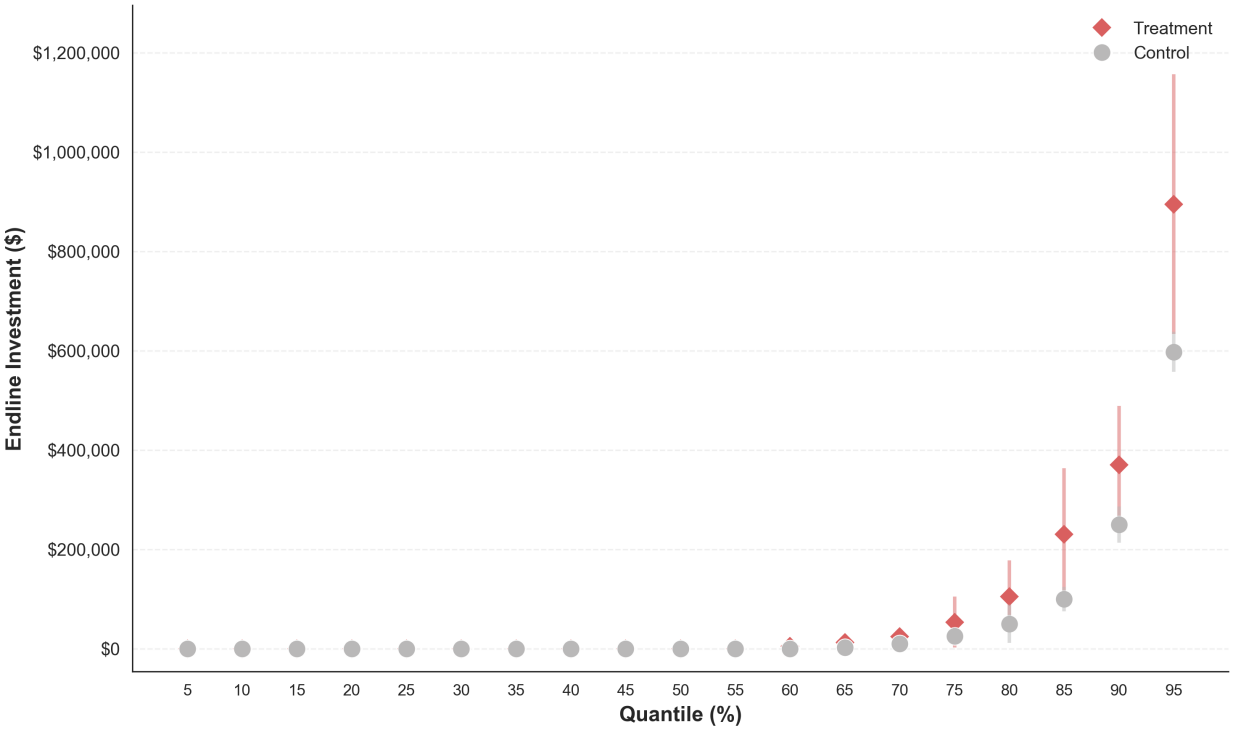
Notes: Panel (a) displays overlapping frequency histograms of cumulative post-baseline tasks completed (weeks 3–10) for treatment and control ventures. Panel (b) reports event study estimates of the treatment effect on cumulative tasks by week, from a panel regression of the outcome on treatment × week interactions, week fixed effects, baseline task count, and strata fixed effects, with standard errors clustered at the venture level. The dashed vertical line marks the start of treatment (after week 2). The shaded region indicates weeks 8–9, during which no weekly surveys were administered; week 10 corresponds to the endline survey. Vertical bars denote 95% confidence intervals. $N = 507$ ventures for distribution of tasks; $N = 515$ for tasks over time.

Figure 8: Revenues by Quantile: Treatment vs. Control



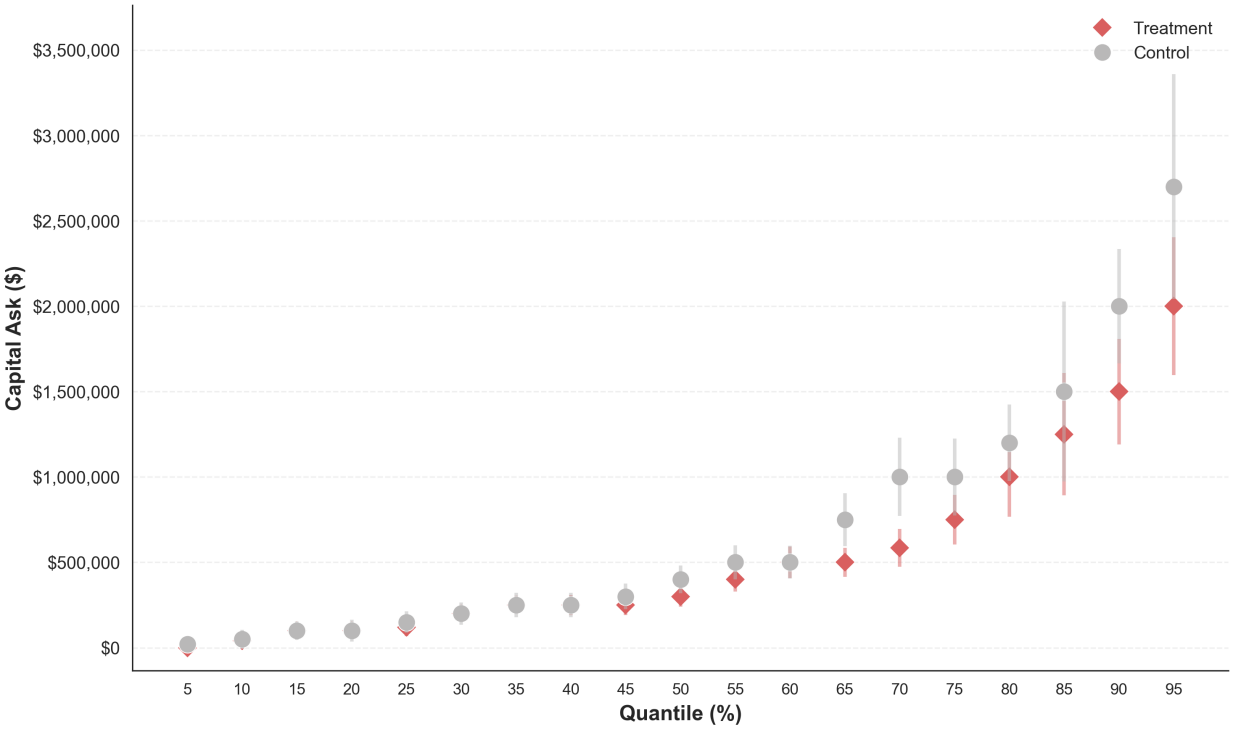
Notes: This figure plots quantile levels of total revenue at Week 10 for treatment and control ventures separately, estimated at every 5th percentile from the 5th to the 95th. Point estimates are sample quantiles computed separately for control and treatment firms. Vertical bars denote 95% confidence intervals from quantile regressions with an intercept only. $N = 479$ ventures.

Figure 9: Investments by Quantile: Treatment vs. Control



Notes: This figure plots quantile levels of total investment raised at Week 10 for treatment and control ventures separately, estimated at every 5th percentile from the 5th to the 95th. Point estimates are sample quantiles computed separately for control and treatment firms. Vertical bars denote 95% confidence intervals from quantile regressions with an intercept only. $N = 479$ ventures.

Figure 10: Capital Demand by Quantile: Treatment vs. Control



Notes: This figure plots quantile levels of self-reported capital demand (capital ask) at Week 10 for treatment and control ventures separately, estimated at every 5th percentile from the 5th to the 95th. Point estimates are sample quantiles computed separately for control and treatment firms. Vertical bars denote 95% confidence intervals from quantile regressions with an intercept only. $N = 479$ ventures.

7 Appendix

This appendix provides supplementary materials organized as follows. We first present evidence on the absence of treatment spillovers (7.1). We then describe our coding procedures for AI use cases (7.2.1) and tasks (7.2.2), and the topic analysis used to categorize AI use cases (7.3). Finally, we present supplementary tables and figures referenced in the main text.

7.1 Accounting for Spillovers

In this section, we examine the possibility of treatment access spillovers, where control group founders might gain access to treatment content, potentially contaminating our estimates. We consider this unlikely for three reasons: (1) we implemented strict design-based safeguards to prevent interaction between treatment and control groups, (2) survey responses show no evidence that control founders were aware of treatment-specific sessions, and (3) attendance records confirm that control founders did not attend treatment-only sessions.

7.1.1 Design-Based Prevention

Throughout the AI Founder Sprint, we took deliberate steps to prevent information spillovers between treatment and control groups. All founders participated through a centralized online platform where we managed announcements, shared resources, and facilitated discussion. Within this platform, treatment and control founders were assigned to completely separate groups with no ability to interact with each other or view each other's content. For Zoom sessions attended by both groups—such as master classes and technical tutorials—we disabled chat functionality to prevent cross-group communication during live sessions.

7.1.2 Survey Evidence

To verify that control founders did not become aware of treatment-specific content, we included a question in the Week 10 endline survey: *"Were there any sessions that you were not able to attend, but you wish you could have? Please be specific and mention who you wanted to learn from."* This question was asked only at endline because founders needed exposure to the full program before they could meaningfully reflect on missed opportunities.

We manually reviewed all 260 responses from the control group founders and categorized them as follows:

- **53 founders (20%)** left the question blank, indicating no additional sessions of interest
- **95 founders (37%)** explicitly stated they attended all sessions or caught up using recordings (e.g., 'Not really,' 'No, attended all the sessions,' 'I attended all')
- **88 founders (34%)** mentioned specific sessions they missed, but all were common sessions offered to both groups, such as Venture Guide meetings, technical tutorials, masterclasses, or

partner sessions. Examples include: 'I wasn't able to attend Grant Lee's session and I really wish I had. His work on storytelling and narrative building would've been so valuable for shaping how we communicate [our] impact..' 'I was not able to attend Masterclass: Claudia Zeisberger and I heard good feedback about it.'

- **24 founders (9%)** gave responses with no evidence of spillover but that did not fit the above categories. Specifically, 14 gave responses too vague to classify (e.g., 'many due to tight schedule'), and 10 offered general program feedback rather than naming a specific missed session (e.g. 'Since my start-up is more in the *impact* space, I would have liked to hear from at least one thought-leader from the Impact Investing space.')

Critically, zero control founders mentioned sessions that were offered exclusively to the treatment group.

7.1.3 Attendance Verification

As a final check for potential treatment access spillovers, we audited attendance records from all treatment-exclusive sessions offered during the AI Founder Sprint.

Attendance verification was conducted using raw Zoom attendance logs. Two research assistants independently coded attendance by session type. Each RA was instructed to match Zoom participant names to founders in the research sample using best-judgment name matching. Because names in Zoom logs can differ slightly from enrollment records (e.g., initials, alternate spellings, or inclusion of startup names), RAs manually reviewed the names to determine who attended which sessions. Attendance was coded as a binary indicator equal to one if the founder appeared in the Zoom logs for a given session at least once, and zero otherwise.

In cases where a participant name was ambiguous—such as when multiple founders shared the same first name—RAs expanded the matching set to include team members, given that team members sometimes attended sessions on behalf of the founder. In all cases, attendance was assigned to the founder or team member in the section corresponding to the session attended. This reflects a conservative and reasonable assumption, given that sessions were section-specific and founders were only provided access to sessions associated with their assigned group.

Based on this coding, no control group founders were found to have attended any treatment-only sessions.

7.2 Coding Decisions for Human Coders

7.2.1 Coding of AI Use Cases

To validate and categorize reported AI use cases, we employed two independent human coders to review all text responses in the `use_case_text` column. These responses came from the survey question: "Did you try out any new ways of using AI in your startup this week? You should include new use cases in your team, as well as new use cases in your product." Coders reviewed each response alongside its associated `venture_id` (unique venture identifier) and `timepoint` (indicating whether the response was submitted during the application phase, week 1, week 2, etc.). For each response, coders filled in two columns: `valid` (coded as 1 for valid, 0 for invalid, or DUP for duplicate) and `category` (coded as product, process, or business model). In cases where coders disagreed, a third researcher adjudicated the disagreement.

Validity Coding Coders assessed whether each text entry constituted a legitimate AI use case. Responses were coded as valid (1) if they described a specific AI application or tool being used for their business, even if relatively vague or broad (e.g., "research," "content creation," "coding"). Examples of valid use cases include:

- "N8N workflow automation for career encyclopedia write-up first draft creation"
- "We have started to use Claude Code and MCP across our software team to boost deployment and testing times"
- "I experimented with Nano Banana and managed to create eCommerce product images"

Responses were coded as invalid (0) if they were completely off-topic, consisted of general comments about AI without specific use cases (e.g., "Yeah, AI is pretty good in creating estimate"), or were nonsense/typos. Explicit "no" or "none" responses were also coded as invalid.

Responses were coded as duplicate (DUP) if the same use case had already been listed by the venture in a previous week. Since our survey explicitly asked founders to report only new AI use cases each week, entries that repeated previously reported use cases—even if not exact matches—were marked as duplicates. For instance, if a venture reported "Claude for coding" in week 1 and "Claude for vibecoding" in week 3, the week 3 entry was coded as a duplicate.

7.2.2 Coding of Tasks

To validate and categorize the reported tasks, two independent human coders reviewed all founder text responses describing the tasks they completed. These responses came from the survey question: "Over the past two weeks, what actions did you take to move your startup forward? Please list all activities, one by one." Coders reviewed each response alongside its associated `venture_id` and `timepoint`. For each response, coders filled in two columns: `valid` (coded as 1 for valid, 0 for invalid, or DUP for duplicate) and `category` (coded as internal or external). Disagreements between coders were resolved by a third researcher.

Validity Coding Coders assessed whether each text entry described a specific task completed to advance the venture. Responses were coded as valid (1) if they described specific business activities—even if brief—such as customer meetings, product development, testing, hiring, fundraising steps, or prototyping. Broad activities like "research" or "troubleshooting" were acceptable as long as they represented concrete actions. Examples of valid tasks include:

- "Talked to 3 clients from Poland (2 Zoom, 1 on-site) to validate pain in booking → invoicing"
- "Demo the platform to Nigerian Union of journalist with over 5000 Members"
- "Launched a $200 Meta Ad Campaign with the service we propose"
- "Onboarded about 10 software testers from the Nigerian software testers association and got feedback from them"

Responses were coded as invalid (0) if they contained meta-references (e.g., "See week 2," "Same as above"), were too vague to identify a concrete action (e.g., "making our product better"), consisted of incomplete fragments, or represented explicit "no/none" responses missed by automated cleaning.

Responses were coded as duplicate (DUP) only when the same activity was listed multiple times with near-exact matches. Coders were instructed to be conservative with duplicate coding, as similar-sounding activities might represent distinct actions. For example, "more sales calls" listed across multiple weeks would be coded as duplicate, but "onboarded 4 more schools" in week 3 and "onboarded 2 more schools" in week 4 would remain distinct, as they clearly refer to different schools.

Category Coding For valid tasks, coders categorized each as either internal or external:

Category	Description	Examples
Internal	Any activity performed within the venture without direct interaction with external people, mostly focused on building, designing, analyzing, or researching internally.	• "Developed MVP of the app" • "Built landing page" • "Coding the platform" • "Created financial projections"
External	Any activity involving people, organizations, or resources outside the startup, mostly related to testing, validating, marketing, or selling the company's idea, product, or business.	• "Had meetings with 3 potential customers to validate product-market fit" • "Pitched to 2 angel investors" • "Launched beta with 10 external testers and collected feedback" • "Sent cold emails to 100 prospects"

7.3 AI Use Cases Topic Analysis

To understand how ventures applied AI across different business functions, we categorized all reported AI use cases from weeks 3–10 (the post-baseline treatment period). The final classification scheme consisted of 7 categories, as shown below, applied to a sample of 3,752 valid post-treatment use cases.

Category	Description	Examples
Product and Strategy Design (n=741, 19.7%)	AI integrated directly into the product as a core feature (e.g., recommendation engines, chatbots, AI-powered functionality) or used to inform high-level strategic decisions such as financial modeling, forecasting, and competitive intelligence.	• "We've begun using AI to run 'what-if' scenarios for our business model shift. We can now simulate the potential outcomes of focusing on the B2C home market versus the B2B school market to better inform our long-term strategy." • "used AI-driven modeling to project cash flow, revenue growth, and hiring timelines, improving decision-making for our pre-seed fundraising strategy." • "Integrated a Gemini-powered LLM into the GST Compliance Module's reconciliation flow. The AI acts as a 'Tax Buddy,' allowing merchants to ask complex, unstructured questions about specific transactions or HSN/SAC codes in natural language, receiving instant, legally-grounded answers instead of navigating dense tax forms." • "Developing an autonomous booking agent for trains: Since most train systems in Africa lack APIs, our AI now detects when a route involves rail and automatically fires up an agent that navigates the train website, fills booking forms, and completes the reservation."

Category	Description	Examples
Research, Legal, and Writing (n=705, 18.8%)	AI used for market research, competitor analysis, tool exploration, grant writing, legal document preparation, compliance review, and knowledge documentation.	• "Used an LLM to help devise a contract that addresses an IP issue we identified" • "Perplexity 'Comet' browser (process) — Used for fast research synthesis to inform proposal writing and use-case scoping" • "Used AI-powered market research tools to analyze competitor offerings and pricing in African digital health"
Product Development (n=681, 18.2%)	AI used for prototyping, coding, debugging, feature development, MVP creation, and software testing.	• "The developer team is trying hands on vibe coding to expedite the deliveries" • "Prototyped an AI event idea generator that creates customized social experiences based on user prompts" • "Using Lovable to design trader screens"
Marketing (n=583, 15.5%)	AI used for content creation, social media management, ad generation, email campaigns, copywriting, and brand communications.	• "Veo 3 to generate an explainer video for our website" • "Used AI-assisted content drafting (problem/solution/business model sections) to prepare accelerator application responses" • "Google Gemini (via NanoBanana) - creative experiments for branding, including early iterations of the [venture] logo"

Category	Description	Examples
Technical Infrastructure (n=488, 13.0%)	AI used for backend systems, infrastructure setup, API integration, cloud deployment, database management, and technical architecture.	• "RAG implementation and improvement for ai response and cost" • "AI for backend APIs integration in Firebase studio" • "We have started to use Claude Code and MCP across our software team to boost deployment and testing times"
Business Operations (n=328, 8.7%)	AI used for workflow automation, process optimization, administrative tasks, HR functions, and operational efficiency.	• "Prototyped an email monitoring workflow with Supabase and n8n that uses AI to categorize messages" • "I used AI to work through my task list of operating the firm" • "Translating the documents from Romanian to Swiss"
Sales (n=226, 6.0%)	AI used for lead generation, sales outreach, customer relationship management, pitch preparation, and sales process optimization.	• "After a discovery call, dropped raw notes into the system and had AI extract contact details, service needs, and urgency level, then automatically populate CRM fields" • "Cluely for the sales calls" • "Used Manus AI to create a LinkedIn Assistant agent"

Categorization Method We developed the taxonomy iteratively using an LLM (Claude Sonnet 4.5 via Claude.ai), combined with keyword-based classification and human-in-the-loop refinement. The taxonomy was derived from a larger dataset combining application-phase use cases with all weekly submissions from Weeks 1–10. After basic data cleaning (removing invalid responses such as 'none' and duplicate entries), this dataset consisted of 7,231 unique use cases. The final classification was then applied to the 4,023 post-baseline use cases used in the analysis.

The process involved multiple rounds of LLM-assisted category generation, keyword expansion, and manual review of uncategorized cases. Through this process, we arrived at an initial set of 12 substantive

categories. After establishing the taxonomy, we classified all 4,023 post-baseline use cases using keyword matching, with ambiguous cases flagged for human review. These reviews helped establish consistent classification rules. For example, we distinguished between use cases referring to features built for external users (classified as *AI as core product feature*) and those describing internal workflow improvements (classified as *Operations and workflow automation*). When multiple classifications were plausible, cases were assigned based on their primary purpose.

In total, 705 cases were manually reviewed. Insights from these decisions were incorporated into an expanded keyword library, and the automated classification was iteratively re-run until all remaining cases were resolved. Cases identified as invalid by human coders were excluded, yielding a final analytical sample of 3,752 use cases.

Following classification, we consolidated the initial 12 categories by merging conceptually similar ones. Tool Evaluation & Exploration, Research & Discovery, Legal Compliance & Regulatory, Documentation & Knowledge Management, and Grant Writing & Funding Applications were combined into *Research, Legal, and Writing*, reflecting their shared focus on knowledge gathering and documentation. AI Product Feature and Data Analytics & Strategic Planning were collapsed into *Product and Strategy Design*, reflecting their shared focus on high-level venture direction and product vision.

7.4 Appendix Tables & Figures

Table A1: Survey Attrition and Panel Completion

Week/Category	All	Control	Treatment	Difference	P-value
WEEKLY RESPONSE RATES					
Week 1	515/515 (100.0%)	260/260 (100.0%)	255/255 (100.0%)	+0.0pp	—
Week 2	515/515 (100.0%)	260/260 (100.0%)	255/255 (100.0%)	+0.0pp	—
Week 3	500/515 (97.1%)	254/260 (97.7%)	246/255 (96.5%)	-1.2pp	0.5739
Week 4	480/515 (93.2%)	244/260 (93.8%)	236/255 (92.5%)	-1.3pp	0.6820
Week 5	453/515 (88.0%)	227/260 (87.3%)	226/255 (88.6%)	+1.3pp	0.7454
Week 6	450/515 (87.4%)	225/260 (86.5%)	225/255 (88.2%)	+1.7pp	0.6548
Week 7	455/515 (88.3%)	232/260 (89.2%)	223/255 (87.5%)	-1.8pp	0.6227
Week 10	479/515 (93.0%)	243/260 (93.5%)	236/255 (92.5%)	-0.9pp	0.8156
ATTRITION PATTERNS					
No missing weeks	417/515 (81.0%)	211/260 (81.2%)	206/255 (80.8%)	-0.4pp	1.0000
1 missing week	31/515 (6.0%)	14/260 (5.4%)	17/255 (6.7%)	+1.3pp	0.6699
2 missing weeks	18/515 (3.5%)	11/260 (4.2%)	7/255 (2.7%)	-1.5pp	0.4978
3 missing weeks	15/515 (2.9%)	7/260 (2.7%)	8/255 (3.1%)	+0.4pp	0.9696
4 missing weeks	17/515 (3.3%)	10/260 (3.8%)	7/255 (2.7%)	-1.1pp	0.6508
5 missing weeks	9/515 (1.7%)	4/260 (1.5%)	5/255 (2.0%)	+0.4pp	0.9766
6 missing weeks	8/515 (1.6%)	3/260 (1.2%)	5/255 (2.0%)	+0.8pp	0.7010
7 missing weeks	0/515 (0.0%)	0/260 (0.0%)	0/255 (0.0%)	+0.0pp	—
8 missing weeks	0/515 (0.0%)	0/260 (0.0%)	0/255 (0.0%)	+0.0pp	—

Table A2: Treatment Effects on Additional Outcomes

	Product Change (1)	Customer Change (2)	Any Pivot (3)	Shut Down (4)
Treatment	0.089***	0.144***	0.057	-0.005
	(0.033)	(0.036)	(0.045)	(0.011)
Strata FE	Yes	Yes	Yes	Yes
Observations	507	507	507	481
Control Mean	0.370	0.226	0.436	0.020
R^2	0.455	0.269	0.055	0.087

Notes: Standard errors are reported in parentheses. Product Change and Customer Change are binary indicators equal to one if the venture moved from having no product to having a product, or from having no customers to having customers, respectively, at any point during the post-baseline period, and zero otherwise. Any Pivot is a binary indicator equal to one if the venture pivoted at least once during the post-baseline period, and zero otherwise. Shut Down is a binary indicator measured at endline; the sample size is 481 because this outcome is constructed using endline survey responses (479 ventures) plus two ventures that responded via email. All specifications include strata fixed effects.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A3: Robustness of Treatment Effects on Total Revenue (Endline)

	Treatment Coef. (SE)	N	R^2	Control Mean
PANEL A: BASELINE TRANSFORMATIONS				
Raw	−1,747,177	479	0.038	2,101,176
	(1,861,794)			
Log (drop zeros)	0.783**	234	0.315	8.950
	(0.347)			
Log$(x+1)$ *[Main Table]*	0.653*	479	0.377	4.171
	(0.382)			
IHS	0.680*	479	0.375	4.486
	(0.407)			
PANEL B: WINSORIZATION ONLY				
Winsorized at 90th percentile	3,728	479	0.313	16,205
	(2,837)			
Winsorized at 95th percentile	10,626	479	0.278	29,765
	(7,265)			
Winsorized at 99th percentile	44,048*	479	0.125	50,279
	(25,491)			
PANEL C: WINSORIZATION AND TRANSFORMATION				
W90 + Log$(x+1)$	0.594	479	0.375	4.049
	(0.364)			
W90 + IHS	0.621	479	0.372	4.364
	(0.389)			
W95 + Log$(x+1)$	0.626*	479	0.379	4.121
	(0.373)			
W95 + IHS	0.652	479	0.377	4.436
	(0.398)			
W99 + Log$(x+1)$	0.667*	479	0.381	4.148
	(0.379)			
W99 + IHS	0.693*	479	0.378	4.463
	(0.404)			

Notes: Each row reports the treatment effect from a separate OLS regression using the indicated transformation of total revenue measured at endline. All specifications include strata fixed effects only. For log specifications that drop zeros, observations with zero revenue at endline are excluded. Control means are reported in the scale of the dependent variable. Standard errors are reported in parentheses.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A4: Robustness of Treatment Effects on Total Investment (Endline)

	Treatment Coef. (SE)	N	R^2	Control Mean
PANEL A: BASELINE TRANSFORMATIONS				
Raw	120,011	479	0.179	107,902
	(89,896)			
Log (drop zeros)	0.572**	195	0.394	10.727
	(0.278)			
Log$(x+1)$ *[Main Table]*	0.436	479	0.280	4.152
	(0.458)			
IHS	0.451	479	0.277	4.418
	(0.486)			
PANEL B: WINSORIZATION ONLY				
Winsorized at 90th percentile	12,265	479	0.364	46,142
	(7,965)			
Winsorized at 95th percentile	16,287	479	0.379	67,066
	(12,871)			
Winsorized at 99th percentile	21,747	479	0.313	93,229
	(23,741)			
PANEL C: WINSORIZATION AND TRANSFORMATION				
W90 + Log$(x+1)$	0.407	479	0.267	4.075
	(0.450)			
W90 + IHS	0.422	479	0.265	4.341
	(0.478)			
W95 + Log$(x+1)$	0.417	479	0.273	4.122
	(0.455)			
W95 + IHS	0.432	479	0.270	4.387
	(0.483)			
W99 + Log$(x+1)$	0.423	479	0.277	4.147
	(0.458)			
W99 + IHS	0.438	479	0.274	4.413
	(0.486)			

Notes: Each row reports the treatment effect from a separate OLS regression using the indicated transformation of total investment measured at endline. All specifications include strata fixed effects only. For log specifications that drop zeros, observations with zero investment at endline are excluded. Control means are reported in the scale of the dependent variable. Standard errors are reported in parentheses.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A5: Robustness of Treatment Effects on Capital Ask (Endline)

	Treatment Coef. (SE)	N	R^2	Control Mean
PANEL A: BASELINE TRANSFORMATIONS				
Raw *[Main Table]*	−223,954**	479	0.094	850,233
	(108,052)			
Log (drop zeros)	-0.089	458	0.167	12.885
	(0.111)			
Log$(x+1)$ *[Main Table]*	-0.502**	479	0.100	12.514
	(0.254)			
IHS	-0.524**	479	0.099	13.187
	(0.265)			
PANEL B: WINSORIZATION ONLY				
Winsorized at 90th percentile	−94,219*	479	0.164	668,340
	(54,379)			
Winsorized at 95th percentile	−107,076*	479	0.156	703,319
	(60,464)			
Winsorized at 99th percentile	−149,198*	479	0.138	780,274
	(78,108)			
PANEL C: WINSORIZATION AND TRANSFORMATION				
W90 + Log$(x+1)$	-0.474*	479	0.097	12.467
	(0.251)			
W90 + IHS	-0.497*	479	0.096	13.140
	(0.263)			
W95 + Log$(x+1)$	-0.480*	479	0.098	12.483
	(0.252)			
W95 + IHS	-0.503*	479	0.097	13.156
	(0.263)			
W99 + Log$(x+1)$	-0.492*	479	0.100	12.505
	(0.253)			
W99 + IHS	-0.515*	479	0.099	13.179
	(0.265)			

Notes: Each row reports the treatment effect from a separate OLS regression using the indicated transformation of capital ask measured at endline. All specifications include strata fixed effects only. For log specifications that drop zeros, observations with zero capital ask at endline are excluded. Control means are reported in the scale of the dependent variable. Standard errors are reported in parentheses.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A6: Heterogeneous Treatment Effects by Founder Technical Background

	Panel A: AI Usage & Tasks Done			
	(1) Number of AI Use Cases	(2) Number of Tasks Done	(3) Number of Internal Tasks	(4) Number of External Tasks
Treatment (Non-Technical Founders)	3.200***	3.379**	2.990***	0.389
	(0.640)	(1.500)	(0.986)	(0.888)
Technical Founder	1.198*	0.809	1.300	-0.491
	(0.685)	(1.428)	(0.949)	(0.864)
Treatment × Technical Founder	-0.997	-2.244	-1.563	-0.681
	(1.032)	(2.067)	(1.349)	(1.247)
Observations	507	507	507	507
Control Mean (Non-Technical)	5.595	18.373	9.302	9.071
R^2	0.152	0.103	0.103	0.089

	Panel B: Venture Performance				
	(5) Venture Progress Index	(6) Product Launched	(7) Has Customers	(8) Has Revenue	(9) Has Investment
Treatment (Non-Technical Founders)	0.300	0.027	0.070	-0.024	0.049
	(0.288)	(0.037)	(0.057)	(0.054)	(0.056)
Technical Founder	-0.171	0.008	-0.052	-0.029	-0.014
	(0.301)	(0.041)	(0.060)	(0.055)	(0.053)
Treatment × Technical Founder	0.258	0.042	0.080	-0.008	-0.021
	(0.410)	(0.051)	(0.083)	(0.079)	(0.078)
Observations	479	479	479	479	479
Control Mean (Non-Technical)	-0.114	0.882	0.672	0.513	0.294
R^2	0.372	0.134	0.176	0.343	0.284

Notes: Standard errors are reported in parentheses. Each column reports estimates from a regression of the outcome on Treatment, an indicator for having a technical founder, and their interaction. Treatment (Non-Technical Founders) reports the treatment effect for ventures without a technical founder. Technical Founder captures the baseline difference between ventures with and without a technical founder in the control group. Treatment × Technical Founder reports the differential treatment effect for ventures with a technical founder relative to those without. Number of AI Use Cases and Number of Tasks Done, Number of Internal Tasks, and Number of External Tasks are cumulative measures over the post-baseline period. The Venture Progress Index is a standardized composite measure aggregating four binary indicators—Product Launched, Has Customers, Has Revenue, and Has Investment—all measured at endline. All regressions include strata fixed effects.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A7: Heterogeneous Treatment Effects by Baseline Performance

	Panel A: AI Usage & Tasks Done			
	(1) Number of AI Use Cases	(2) Number of Tasks Done	(3) Number of Internal Tasks	(4) Number of External Tasks
Treatment (Low Traction)	2.901***	2.305*	2.446***	-0.140
	(0.638)	(1.315)	(0.864)	(0.783)
High Traction	0.318	2.485	0.118	2.367*
	(1.189)	(2.450)	(1.696)	(1.296)
Treatment × High Traction	-0.581	-0.079	-0.662	0.583
	(1.000)	(2.117)	(1.355)	(1.341)
Observations	507	507	507	507
Control Mean (Low Traction)	6.155	18.820	10.205	8.615
R^2	0.148	0.101	0.100	0.086

	Panel B: Venture Performance				
	(5) Venture Progress Index	(6) Product Launched	(7) Has Customers	(8) Has Revenue	(9) Has Investment
Treatment (Low Traction)	0.614**	0.061	0.174***	-0.032	0.043
	(0.272)	(0.037)	(0.056)	(0.051)	(0.046)
High Traction	4.290***	0.142*	0.629***	0.680***	0.509***
	(0.467)	(0.084)	(0.105)	(0.080)	(0.067)
Treatment × High Traction	-0.514	-0.039	-0.178**	0.012	-0.009
	(0.399)	(0.042)	(0.076)	(0.077)	(0.083)
Observations	479	479	479	479	479
Control Mean (Low Traction)	-1.378	0.837	0.484	0.281	0.183
R^2	0.374	0.131	0.182	0.342	0.284

Notes: Standard errors are reported in parentheses. Each column reports estimates from a regression of the outcome on Treatment, an indicator for high baseline venture performance, and their interaction. Treatment (Low Traction) reports the treatment effect for low-traction ventures. High Traction captures the baseline difference between high- and low-traction ventures in the control group. Treatment × High Traction reports the differential treatment effect for high-traction ventures relative to low-traction ventures.

Number of AI Use Cases and Number of Tasks Done, Number of Internal Tasks, and Number of External Tasks are cumulative measures over the post-baseline period. The Venture Progress Index is a standardized composite measure aggregating four binary indicators—Product Launched, Has Customers, Has Revenue, and Has Investment—all measured at endline. All regressions include strata fixed effects.

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A8: Example AI Use Cases from Treatment Ventures

Use Case	Production Shift
"Worked Magic Patterns into our product design process. We've eliminated our part-time designer."	Design role eliminated; AI directly substitutes for specialized hire
gray!30	Expert knowledge role replaced by AI agent; no hire needed
black "We leveraged Manus as our Clinical Affairs Specialist, drawing on its medical/clinical expertise in gynecology, radiology, trial coordination, ethics approvals, and maternal health. We tasked Manus with researching the requirements for conducting a clinical trial and creating an implementation plan. It was delivered successfully, enabling us to proceed with clinical trials without hiring for that role."	
gray!30	Repeated manual inspections replaced by vision model in production
black "We started using in production Uworktify Eye V1, with automatize with AI the car inspection workflow that have to be done for each car several times a day."	
gray!30	Compliance and hiring fully transferred to AI
black "Onboarding and KYC is 100% AI screened and controlled." "We now use AI for new employment interviews."	
gray!30	Four-step human relay replaced by single AI chain
black "Voice-note to scope: ASR + LLM chain turns WhatsApp voice notes into a structured job with exactly one clarifying question when needed."	
gray!30	Clinical intake radically sped up; 1 hr → 8 min
black "We tested an AI chatbot for onboarding that guides mothers through 5–7 health questions via WhatsApp and classifies them as low, moderate, or high risk, reducing onboarding from 1 hour to under 8 minutes."	
gray!30	Human triage queue replaced by instant AI classifier
black "Inbound intent + urgency classifier: zero-shot model routes messages into Book / Quote / Reschedule / Billing / Spam with a confidence score that drives auto-approve vs. review."	
gray!30	Manual data entry replaced by voice capture; +40% collection rate
black "Speech-to-Text Data Collection: Integrated voice input for field agents in open markets through DataX, increasing collection rate by over 40%."	
gray!30	Multi-source report drafting collapsed into single automated pipeline
black "Used Claude Code to complete an end-to-end automated report writer for an education client, which pulls teaching-related information from diverse databases and APIs."	
gray!30	Five-step process reduced to one automated trigger
black "Customer leaves 4-star review → AI analyzes → generates response → sends WhatsApp discount → updates FAQ."	
gray!30	Multi-turn sales conversation compressed to one-step AI lookup
black "Pricebook auto-match + clarifier: Maps messy requests to SKUs and asks a single clarifying question only when confidence is low."	

Notes: Each row reports a use case submitted by a treatment venture in weekly progress reports (weeks 3–10). Quotes are reproduced verbatim. All ventures are in the treatment arm. Venture identifiers anonymized.

Table A9: Illustrative Examples of AI-Driven Production Reorganization in Treatment Ventures

Domain	Existing Process & Constraint	AI Reorganization	Outcome
Field Services	Dispatcher coordinates bookings, quotes, invoicing, and collections via phone, WhatsApp, and email	19 use cases rebuilt entire production chain: intent classifier replaced dispatcher; address/VAT normalizer replaced bookkeeper; slot proposer replaced scheduler; multilingual adapter replaced comms staff; payment-reminder engine replaced collections. Confidence watchdog routes uncertain cases to humans, then learns from corrections.	5 roles compressed into AI modules; system self-improves from operator feedback
gray!30 black Voice Commerce	Literacy and language barriers block kirana shopkeepers from digital adoption; no designer on team	Voice ordering replaces typed interfaces; AI generates storefronts from photos; phonetic disambiguation maps dialect to SKUs ("Dant Kanti" → Dabur Red Toothpaste); GST tax classification and real-time P&L via voice query.	3 retailer back-office roles compressed; founder's missing designer filled by AI
gray!30 black Trade QC	Human inspectors limit cross-border quality verification; trust depends on inspector reputation	Multi-modal pipeline: image forensics detects manipulated photos; LLM drafts compliance reports; ESG tagger extracts sustainability data; location verification cross-checks GPS; confidence scoring rates inspector reliability. Tested with DP World.	70% reduction in manual review; \$2K/mo → \$50K/wk revenue; 2 → 40 customers
gray!30 black Fintech	KYC, call center, and hiring consume labor costs at scale (2,500 existing customers)	100% AI-screened KYC; AI call center (Hume + ElevenLabs + Twilio); AI-conducted hiring interviews; Synthesia training videos for new hires; AI content fills marketing gap.	Customer base +20% while restructuring 3 cost centers from human to AI operations
gray!30 black Maternal Health	Clinical intake requires trained health workers; manual risk scoring is slow and expensive	WhatsApp chatbot conducts intake; AI generates multi-risk scores (clinical, behavioral, socioeconomic); recommendation layer prescribes action plans; predictive model flags likely complications 4 weeks ahead. Quality system learns from clinician corrections.	Onboarding: 1 hr → 8 min; real-time dashboards replace weeks of manual reporting
gray!30 black SME Marketing	Training partners limits scaling; each agency requires hands-on onboarding	AI generates campaigns from single business description; AI creates training materials to teach partners how to use the tool ("robot teaches robot"); tracks employee progress and adapts training. Voice-of-customer mining generates content using customers' own language.	Revenue +62% (\$13K → \$21K/wk); recursive compression across product, training, and marketing
gray!30 black Tender Matching	Manual scanning of procurement tenders is economically unviable for small teams	End-to-end pipeline: classify tenders → match to SMEs → auto-email recommendations → RAG-assisted proposal drafting → compliance check (flags missing ISO docs) → win-probability scoring → predictive bid pricing.	First revenue: \$0 → \$40K; 0 → 4 paying customers; product launched during program
gray!30 black Medical Transport	Human dispatcher limits ride throughput; scheduling is reactive	AI dispatch routes by patient urgency and traffic; demand forecasting pre-positions drivers before discharge peaks; conversational SMS handles booking; automated reconciliation calculates commissions.	160 rides/wk; 86% driver acceptance; 60% bookings via AI; reactive → predictive scheduling

Notes: Each row illustrates a treatment venture describing how they used AI to reorganize production. All ventures are in the treatment arm; use cases drawn from weekly progress reports (weeks 3–10). Venture identifiers anonymized.

Table A10: Quantile Treatment Effects on Revenue (Raw)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	0	0	0.00	479
			(518.84)	
0.10	0	0	0.00	479
			(461.25)	
0.15	0	0	0.00	479
			(473.32)	
0.20	0	0	0.00	479
			(445.21)	
0.25	0	0	0.00	479
			(413.54)	
0.30	0	0	0.00	479
			(435.99)	
0.35	0	0	0.00	479
			(311.37)	
0.40	0	0	0.00	479
			(287.34)	
0.45	0	0	0.00	479
			(250.41)	
0.50	0	120	0.00	479
			(167.15)	
0.55	5.50	552	0.00	479
			(153.00)	
0.60	280	1,500	0.00	479
			(76.38)	
0.65	1,530	5,875	0.00	479
			(114.01)	
0.70	3,840	10,000	0.00	479
			(473.23)	
0.75	7,000	16,025	20.00	479
			(543.49)	
0.80	13,687.20	28,000	328.94	479
			(786.49)	
0.85	44,637.60	59,500	2,380.00	479
			(2,848.43)	
0.90	98,000	151,000	3,500.00	479
			(5,789.74)	
0.95	231,500	455,495.75	22,700.00**	479
			(9,033.81)	

Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: Quantile Treatment Effects on Revenue (W99)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	0	0	0.00	479
			(518.84)	
0.10	0	0	0.00	479
			(461.25)	
0.15	0	0	0.00	479
			(473.32)	
0.20	0	0	0.00	479
			(444.38)	
0.25	0	0	0.00	479
			(413.18)	
0.30	0	0	0.00	479
			(436.01)	
0.35	0	0	0.00	479
			(311.37)	
0.40	0	0	0.00	479
			(286.93)	
0.45	0	0	0.00	479
			(250.41)	
0.50	0	120	0.00	479
			(175.73)	
0.55	5.50	552	0.00	479
			(153.00)	
0.60	280	1,500	0.00	479
			(83.58)	
0.65	1,530	5,875	0.00	479
			(114.01)	
0.70	3,840	10,000	0.00	479
			(472.37)	
0.75	7,000	16,025	20.00	479
			(585.62)	
0.80	13,687.20	28,000	1,081.73	479
			(764.71)	
0.85	44,637.60	59,500	2,380.00	479
			(2,760.51)	
0.90	98,000	151,000	3,500.00	479
			(5,811.76)	
0.95	231,500	455,495.75	22,700.00**	479
			(9,033.87)	

Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the winsorized outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12: Quantile Treatment Effects on Total Investment (Raw)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	0	0	0.00	479
			(1,741.06)	
0.10	0	0	0.00	479
			(1,541.91)	
0.15	0	0	0.00	479
			(1,461.75)	
0.20	0	0	0.00	479
			(1,327.00)	
0.25	0	0	0.00	479
			(1,226.03)	
0.30	0	0	0.00	479
			(1,177.11)	
0.35	0	0	0.00	479
			(1,025.75)	
0.40	0	0	0.00	479
			(953.65)	
0.45	0	0	0.00	479
			(907.29)	
0.50	0	0	0.00	479
			(739.75)	
0.55	0	0	0.00	479
			(584.56)	
0.60	0	5,000	0.00	479
			(792.07)	
0.65	2,150	11,875	0.00	479
			(532.61)	
0.70	10,000	23,750	0.00	479
			(1,211.43)	
0.75	22,500	50,875	0.00	479
			(2,238.64)	
0.80	50,000	105,000	4,000.00	479
			(3,242.01)	
0.85	100,000	225,750	4,000.00	479
			(4,746.42)	
0.90	250,000	360,000	24,945.00**	479
			(10,097.87)	
0.95	592,750	823,750	31,000.00*	479
			(18,150.83)	

Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A13: Quantile Treatment Effects on Total Investment (W99)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	0	0	0.00	479
			(1,741.06)	
0.10	0	0	0.00	479
			(1,541.91)	
0.15	0	0	0.00	479
			(1,461.75)	
0.20	0	0	0.00	479
			(1,322.15)	
0.25	0	0	0.00	479
			(1,227.55)	
0.30	0	0	0.00	479
			(1,177.11)	
0.35	0	0	0.00	479
			(1,025.75)	
0.40	0	0	0.00	479
			(953.61)	
0.45	0	0	0.00	479
			(907.29)	
0.50	0	0	0.00	479
			(746.32)	
0.55	0	0	0.00	479
			(584.56)	
0.60	0	5,000	0.00	479
			(704.11)	
0.65	2,150	11,875	0.00	479
			(532.61)	
0.70	10,000	23,750	0.00	479
			(1,224.84)	
0.75	22,500	50,875	0.00	479
			(2,091.50)	
0.80	50,000	105,000	4,000.00	479
			(3,243.58)	
0.85	100,000	225,750	4,000.00	479
			(4,746.42)	
0.90	250,000	360,000	20,500.77**	479
			(10,349.43)	
0.95	592,750	823,750	31,000.00*	479
			(16,452.95)	

Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the winsorized outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A14: Quantile Treatment Effects on Capital Ask (Raw)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	20,500	0	-5,000.00	479
			(28,188.21)	
0.10	50,000	45,000	-15,000.00	479
			(31,208.16)	
0.15	100,000	100,000	0.00	479
			(32,954.84)	
0.20	100,000	100,000	0.00	479
			(33,453.70)	
0.25	150,000	115,000	0.00	479
			(34,316.51)	
0.30	200,000	200,000	0.00	479
			(35,118.20)	
0.35	250,000	250,000	0.00	479
			(36,032.39)	
0.40	250,000	250,000	0.00	479
			(37,069.49)	
0.45	300,000	250,000	0.00	479
			(37,598.63)	
0.50	400,000	300,000	0.00	479
			(41,909.76)	
0.55	500,000	425,000	0.00	479
			(42,322.63)	
0.60	500,000	500,000	-50,000.00	479
			(46,326.93)	
0.65	750,000	500,000	-50,000.00	479
			(46,394.58)	
0.70	1,000,000	583,500	-146,819.66***	479
			(54,145.90)	
0.75	1,000,000	750,000	-177,059.35***	479
			(62,567.06)	
0.80	1,120,000	1,000,000	-200,000.00***	479
			(71,924.98)	
0.85	1,500,000	1,187,500	0.00	479
			(93,933.95)	
0.90	2,000,000	1,500,000	-250,000.00*	479
			(128,664.88)	
0.95	2,680,000	2,000,000	-500,000.00***	479
			(163,325.75)	

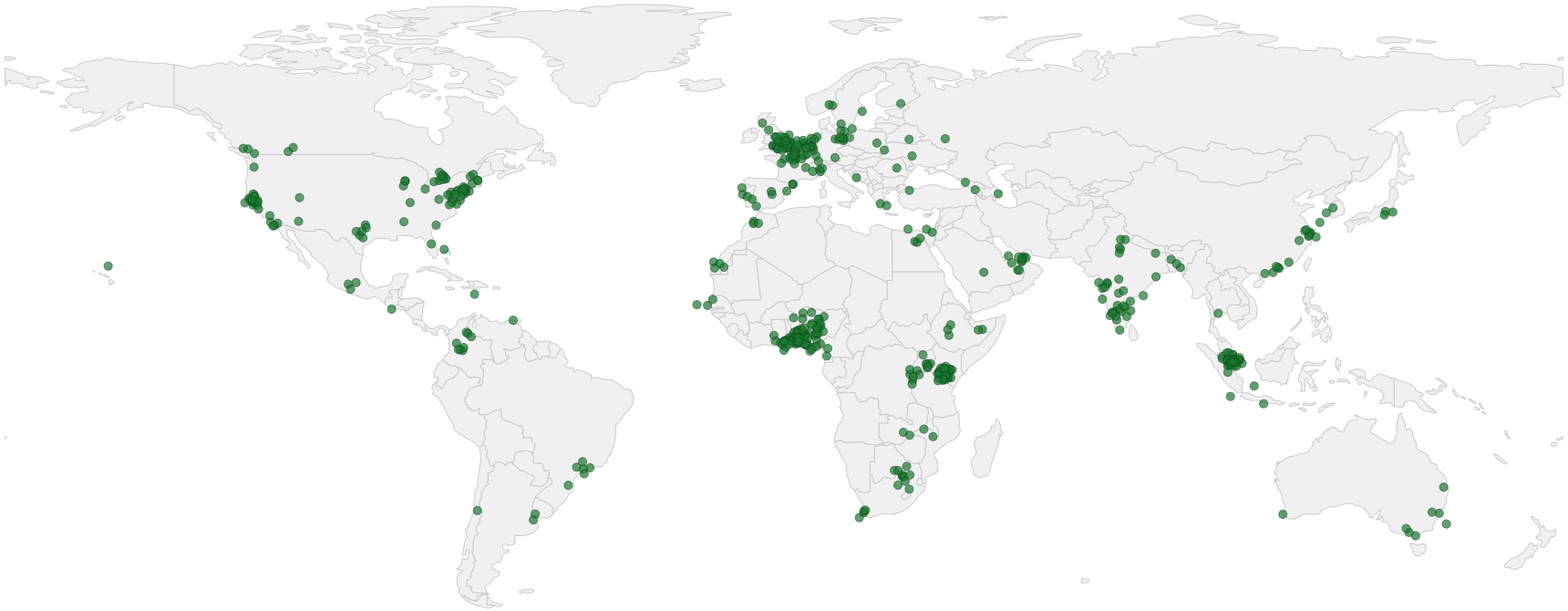
Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A15: Quantile Treatment Effects on Capital Ask (W99)

Quantile	Control q	Treatment q	Treatment effect	N
0.05	20,500	0	-5,000.00	479
			(28,188.21)	
0.10	50,000	45,000	-15,000.00	479
			(30,731.89)	
0.15	100,000	100,000	0.00	479
			(32,954.84)	
0.20	100,000	100,000	0.00	479
			(33,559.36)	
0.25	150,000	115,000	0.00	479
			(34,320.92)	
0.30	200,000	200,000	0.00	479
			(35,089.03)	
0.35	250,000	250,000	0.00	479
			(36,032.39)	
0.40	250,000	250,000	0.00	479
			(37,254.94)	
0.45	300,000	250,000	0.00	479
			(37,598.43)	
0.50	400,000	300,000	0.00	479
			(42,060.21)	
0.55	500,000	425,000	0.00	479
			(42,322.63)	
0.60	500,000	500,000	-50,000.00	479
			(46,350.26)	
0.65	750,000	500,000	-50,000.00	479
			(46,429.94)	
0.70	1,000,000	583,500	-132,798.11**	479
			(53,100.57)	
0.75	1,000,000	750,000	-150,001.06**	479
			(62,777.62)	
0.80	1,120,000	1,000,000	-200,000.00***	479
			(70,332.18)	
0.85	1,500,000	1,187,500	0.00	479
			(93,826.37)	
0.90	2,000,000	1,500,000	-250,000.00*	479
			(128,675.07)	
0.95	2,680,000	2,000,000	-500,000.00***	479
			(156,917.10)	

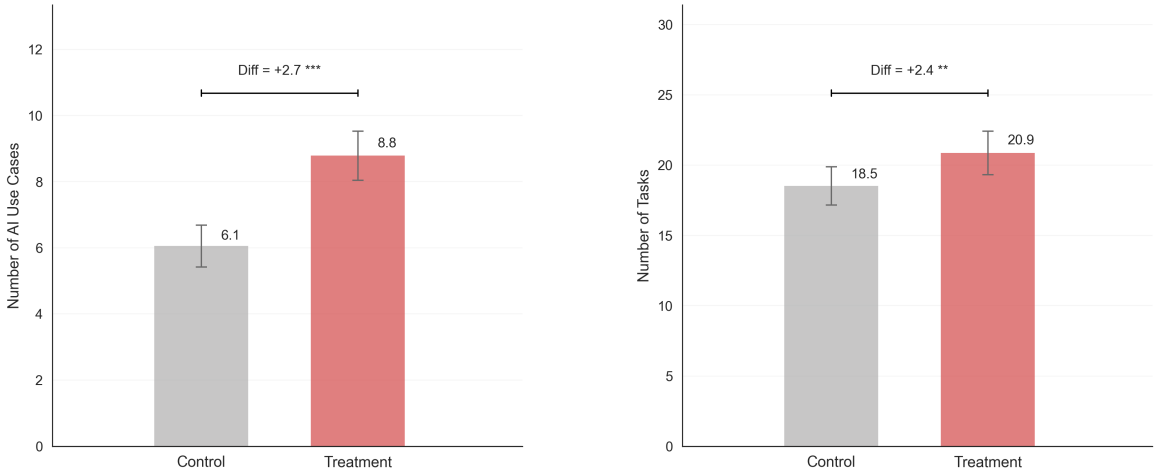
Notes: Each row reports the treatment effect from a separate quantile regression at the indicated quantile. Control q and Treatment q report the corresponding sample quantiles of the winsorized outcome in the control and treatment groups, respectively. All specifications include strata fixed effects. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure A1: Geographic Distribution of Startups in the Experiment



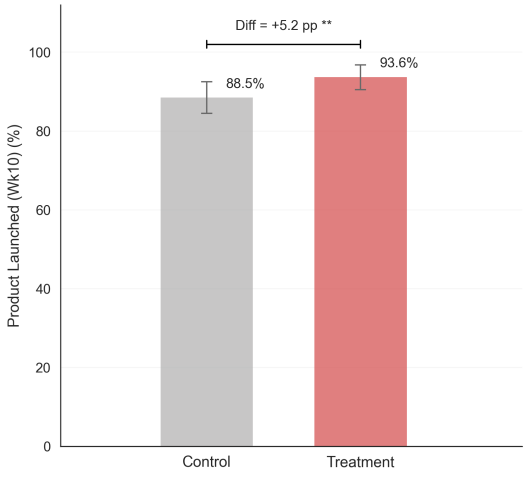
Notes: Each dot represents one of the startups in the experiment, placed at the startup's HQ city with random jitter ($\sigma = 1°$) added to reveal overlapping observations. City coordinates were estimated using Nominatim geocoding for 514 of the 515 startups in our sample.

Figure A2: Treatment effects on key outcomes

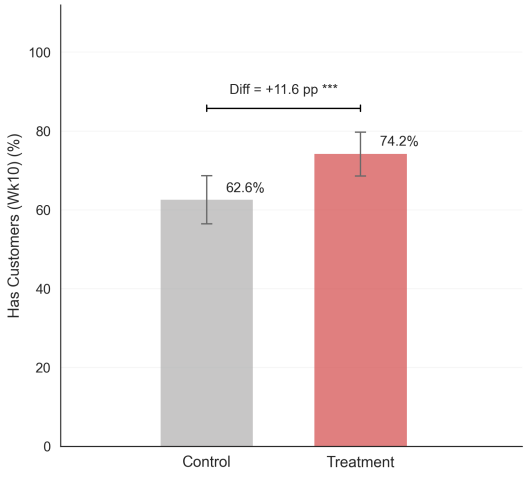


(a) Cumulative AI Use Cases **(b) Cumulative Tasks**

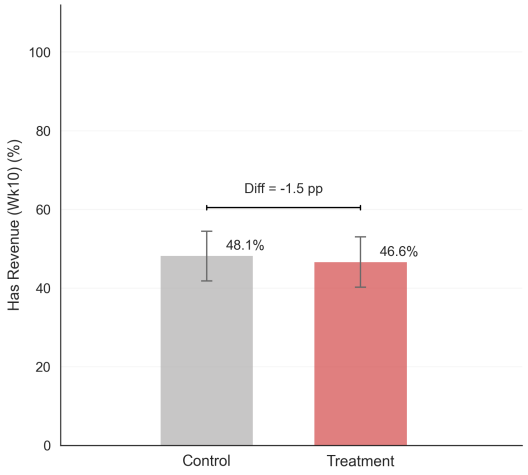
Figure A2: Treatment effects on key outcomes (Continued)



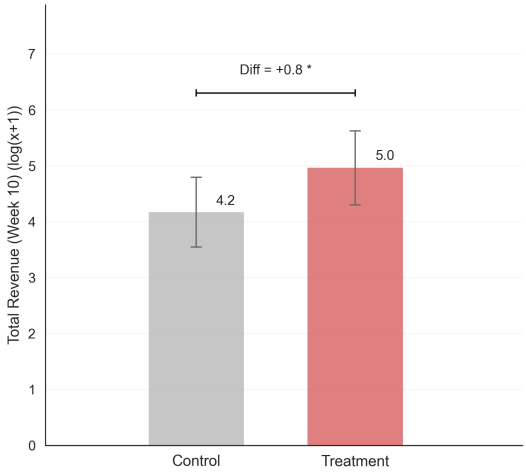
(c) Product Launched



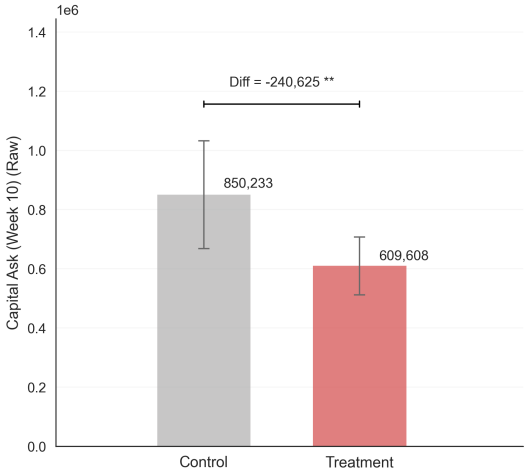
(d) Customers Acquired



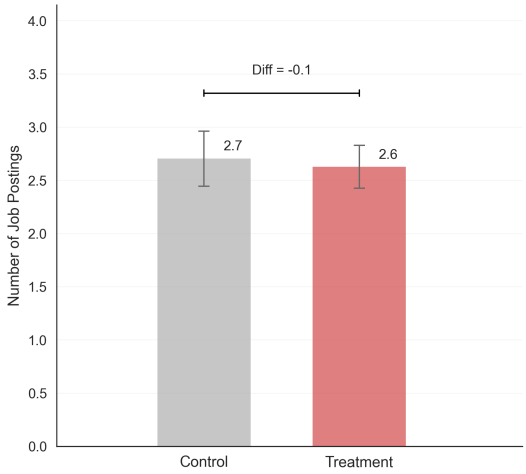
(e) Has Revenue



(f) Total Revenue Generated

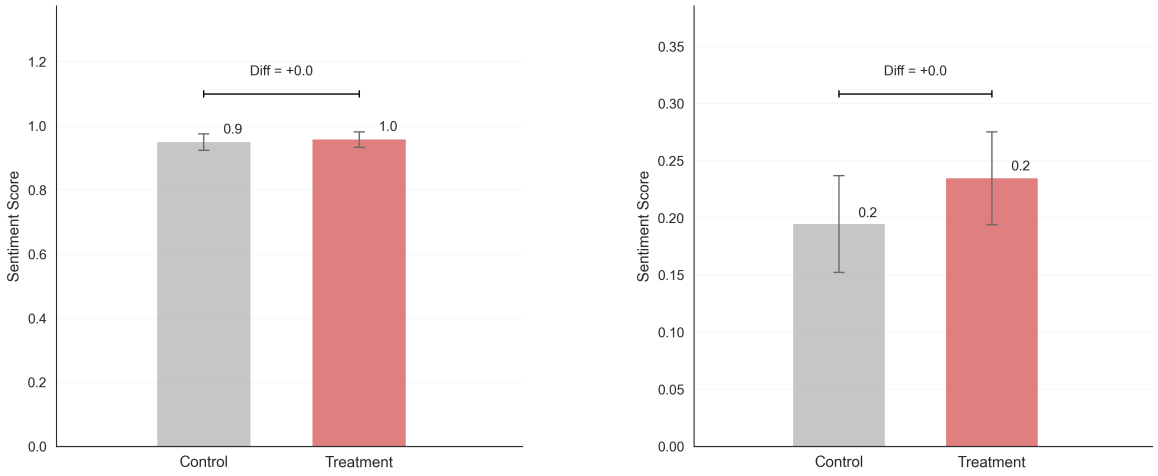


(g) Capital Demand



(h) Labor Demand

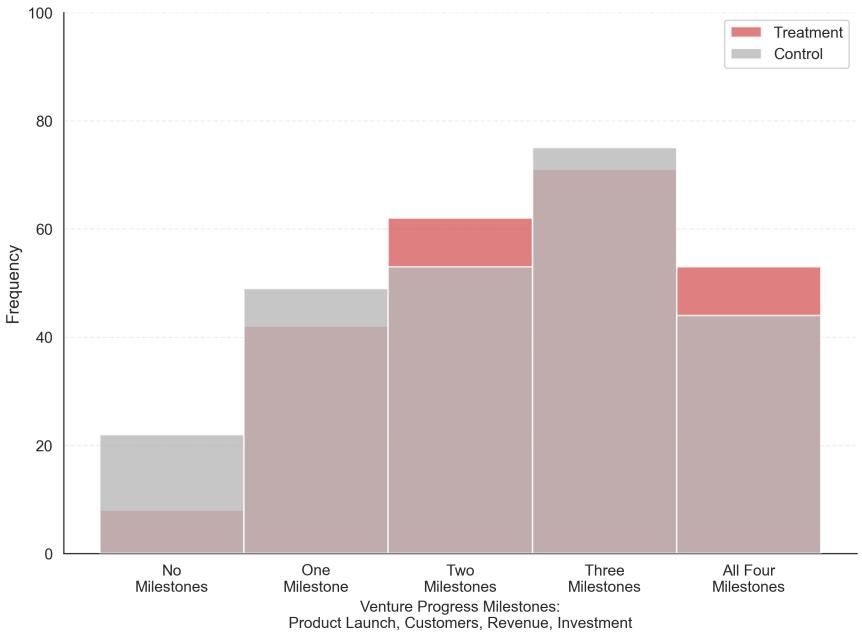
Figure A3: Sentiment analysis results



(a) VADER sentiment **(b) RoBERTa sentiment**

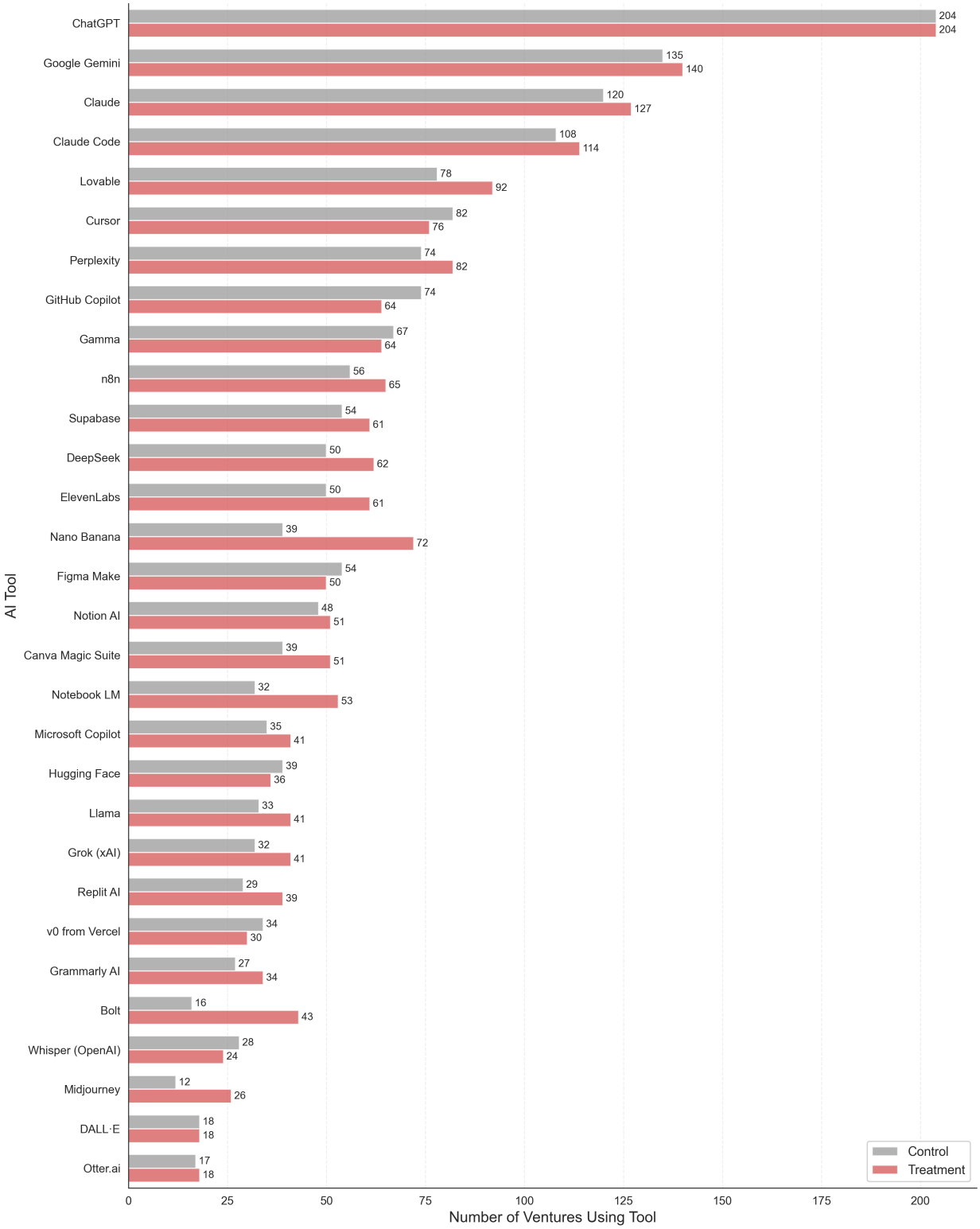
Notes: Sentiment scores are computed from a single consolidated text blob constructed by concatenating multiple post-baseline open-ended text fields. These include all post-baseline AI use case submissions, all post-baseline task submissions, endline venture narrative fields (problem, solution, and business model), and additional endline reflection text. Reflection text includes founders' descriptions of what they found most impactful during the accelerator, which program elements they found most useful and why, why they believe they are best positioned for Demo Day, and any qualitative feedback provided about the accelerator.

Figure A4: Distribution of Venture Progress Milestones



Notes: This figure shows the distribution of the venture progress milestones across treatment and control groups, measured at endline. The index sums four binary milestones: product launched, has customers, has revenue, and has investment (range 0–4). $N = 479$ ventures with non-missing endline data.

Figure A5: 30 Most Frequently Used AI Tools: Treatment vs. Control



Notes: This figure shows the 30 most frequently used AI tools reported by ventures at Week 10. Ventures were presented with a curated list of 50 tools and could select all that applied; those who selected "Other" were given the option to free-text additional tools not included in the list. Both curated selections and free-text responses are reflected in the counts.