

The Uneven Impact of Generative AI on Entrepreneurial Performance: Evidence from a Field Experiment in Kenya

Nicholas G. Otis*
Berkeley Haas

Rowan Clarke
Harvard Business School

Solène Delecourt
Berkeley Haas

David Holtz
Columbia Business School

Rembrand Koning
Harvard Business School

Scalable and low-cost AI assistance has the potential to improve firm decision-making and economic performance, particularly in emerging markets. However, running a business involves a wide range of open-ended problems, making it unclear whether and how recent advances in AI can help business owners around the world make better decisions. In a field experiment with Kenyan entrepreneurs, we evaluated the impact of AI advice on small business revenues and profits by randomizing access to a GPT-4-powered AI business assistant. While we are unable to reject the null hypothesis of no *average* treatment effect on firm revenues and profits, we find that the effect for entrepreneurs who were low-performing at baseline is nearly 0.25 standard deviations lower than for initial high performers. Subsample analyses show that low performers did nearly 10% worse due to the AI assistant, whereas high performers may have benefited by over 15%. This differential impact does not appear to result from differences in the questions posed to or the advice received from the AI, but rather from which pieces of AI advice entrepreneurs selected and implemented. More broadly, these results show that generative AI is already capable of impacting real-world business performance—though in uneven and sometimes unexpected ways.

*Otis: notis@berkeley.edu (corresponding author); Clarke: rclarke@hbs.edu; Delecourt: solened@berkeley.edu; Holtz: david.holtz@columbia.edu; Koning: rem@hbs.edu. All authors after Otis are listed in alphabetical order. The authors gratefully acknowledge financial support from the Agency Fund, the South Park Common Social Impact Fellowship, the Weiss Fund, Microsoft, the Cora Jane Flood Endowment at Berkeley Haas, Berkeley Haas, the Digital Data Design (D³) Institute at Harvard, and Harvard Business School. The authors thank the Busara Center, and especially Chaning Jang for their exceptional support in our field operations, and Sara Dahiya, Brian Mutisyo, Elyes Laalai, Kelsey Shipman, and Noa Ben Haim for superb research assistance. The authors are also grateful to M. Hassan Siddique and M. Tariq Sajid for their assistance in developing the AI assistant, and to Alex Pompe for his assistance with this project. We have benefited from seminar and conference feedback at UC Berkeley, UPenn Wharton, D³, MIT CODE, CFXS, Maryland, Strategy Science, INSEAD, HBS, AEIA, Microsoft, Northeastern, the World Bank, and the University of Washington. This project was approved by the U.C. Berkeley Office for the Protection of Human Subjects.

1 Introduction

Since the launch of ChatGPT in November 2022, there has been an explosion of research on generative AI and its potential economic implications (Mollick, 2022; The White House, 2022; Agrawal, Gans, and Goldfarb, 2023; Eloundou et al., 2024; Goldfarb, 2024). Much of this recent work is driven by the belief that conversations with large language models (LLMs) can help people learn and develop new skills (Choi et al., 2023; Mollick and Mollick, 2022), and in work contexts, can improve firm performance and growth (Brynjolfsson, Li, and Raymond, 2025; Dell’Acqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Kumar et al., 2023). Given the substantial variation in worker and firm productivity, both within and between countries, the emergence of almost zero marginal-cost generative AI “assistants” has the potential to radically improve the productivity and performance of everyone, from the thousands of CEOs running companies listed on the New York Stock Exchange to the hundreds of millions of entrepreneurs running small and medium-sized businesses in developing economies (McAfee, Rock, and Brynjolfsson, 2023; Björkegren, 2023).

Consistent with the optimism currently surrounding generative AI (AI hereafter), recent experiments show that conversing with AI and receiving AI assistance causes workers to write better business text, including press releases, ad copy, consulting memos, and customer support messages, more quickly (Brynjolfsson, Li, and Raymond, 2025; Dell’Acqua et al., 2023; Doshi and Hauser, 2023; Noy and Zhang, 2023; Chen and Chan, 2023). However, it remains unclear how the benefits of such AI feedback generalize to the broader set of tasks that firms engage in. Beyond these text-based tasks, firms must also manage employees, raise capital, pilot new initiatives, run advertising strategies, price their services, react to competitors, and decide which of these and myriad other tasks to focus their efforts on (Chandler, 1977). The sheer multitude of tasks involved in running a business greatly increases the complexity of effectively learning how to improve business performance, especially for small businesses and entrepreneurs in emerging markets, which often lack the team, training, and advisors that drive learning (Kim, 2024; Hanna, Mullainathan, and Schwartzstein, 2014). And even when they have access to human advisors and mentors, entrepreneurs often struggle to decide which tasks they should ask for help with, how to formulate effective questions to get useful

feedback, and how to interpret and act upon the advice they receive (Bryan, Tilesik, and Zhu, 2017; Camuffo et al., 2020; Agrawal, Gans, and Stern, 2021; Dimitriadis and Koning, 2022).

It’s possible that these challenges are exacerbated when seeking advice from AI systems. Entrepreneurs may be reluctant to ask questions of an AI assistant, rather than another person (Lebovitz, Lifshitz-Assaf, and Levina, 2022). Without good questions, even advanced AI systems may fail to provide relevant or useful guidance. And even if an entrepreneur’s question is well-crafted, it is unclear how useful AI advice will be in practice. Given that many business tasks are tacit and may not yet be codified as text—especially tasks in emerging markets—AI tools may lack the relevant training data to provide helpful answers (Autor, 2014; Tao et al., 2023). AI systems can also produce ineffective or overconfident “solutions” that, if implemented, could worsen rather than improve firm performance (Dell’Acqua et al., 2023; Ji et al., 2023), and entrepreneurs may lack the judgment, complementary knowledge, skills, or resources needed to select and implement helpful AI-generated suggestions while avoiding harmful ones (Brynjolfsson and Hitt, 2000; Brynjolfsson, Rock, and Syverson, 2021; Agrawal, Gans, and Stern, 2021). However, if entrepreneurs are able to formulate good questions, if AI provides useful suggestions, and if entrepreneurs can effectively use their judgment to act on this guidance, then AI advice has the potential to help entrepreneurs, especially entrepreneurs in emerging markets.

In this paper, we explore the potential of AI to impact firm performance in a randomized controlled trial (RCT) with 640 Kenyan small business entrepreneurs, half of whom were randomly assigned to receive access to a GPT-4-powered AI assistant over WhatsApp. We are unable to reject the null hypothesis that, on average, access to generative AI had no effect on business performance relative to the control. Although our overall treatment effect estimate is statistically indistinguishable from zero, this result masks substantial treatment effect heterogeneity with respect to pre-treatment firm performance. We estimate that the effect of treatment for entrepreneurs with below-median baseline performance (henceforth “low performers”) is nearly 0.25 standard deviations lower than for entrepreneurs with above-median baseline performance (“high performers”). Subsample analysis shows that receiving the AI assistant led to a nearly 10% decrease in performance for low performers, but that

for high performers receiving the AI assistant may have caused a more than 15% increase in performance relative to the control.

To better understand the mechanisms driving this heterogeneity, we engage in an exploratory text analysis of the questions entrepreneurs asked, the answers generated by the AI assistant, and post-treatment survey responses from the entrepreneurs describing their business changes. This analysis reveals no evidence that the performance effects we observe are rooted in differences in the number or kinds of questions asked, in the types of suggestions generated by the AI, or in the likelihood of making business changes based on the AI's advice. Instead, we find that low and high performers selected and then implemented different aspects of the nearly two dozen suggestions that the average entrepreneur received from our AI assistant. We find that treated low performers were especially likely to implement "generic" advice focused on lowering prices and investing in advertising, which, if inappropriate for the entrepreneur's business situation, could decrease revenues and increase costs. In contrast, treated high performers worked with the AI to discover tailored and specific changes that benefited their businesses. These changes ranged from finding alternative ways to get power during blackouts to introducing a new in-demand car wash detergent. These findings highlight that our AI assistant enabled both low- and high-performing entrepreneurs to ask questions, receive more advice, and act on this AI advice. However, only high performers were able to both effectively screen and implement valuable, as opposed to detrimental, AI-generated suggestions.

Our results make three primary contributions to our understanding of the economic impact of AI. First, understanding whether AI can assist entrepreneurial decision making provides more general insights into whether generative AI can provide meaningful benefits in contexts that are more complex and interconnected than the well-defined tasks studied in previous work (Sato et al., 2023). Extending estimates from research studying well-defined work tasks (Noy and Zhang, 2023; Dell'Acqua et al., 2023), our results suggest that while AI can already produce a plethora of potentially useful advice for a multitude of business tasks, understanding what types of task advice users choose to implement will be crucial to determine who will benefit from these new AI tools (Dell'Acqua et al., 2023; Wiles and Horton, 2024). Second, our research sheds light on the potential impact of AI in

emerging markets, a setting that has been largely overlooked in recent studies of generative AI (Björkegren, 2023). Third, we show that while generative AI offers an opportunity to scale the benefits of business mentorship and training all over the world (McKenzie, 2021), successfully scaling up this new technology to benefit everyone will require improvements to both the underlying technology and the way we train entrepreneurs and workers to use these tools.

More generally, our study highlights the vast number of contingencies at play when assessing the economic impacts of AI in different contexts. Would our AI tool have a similar uneven impact for entrepreneurs in the US who on average have more education and experience? How might our results change if the entrepreneurs in our study only used AI for a well-defined task, like writing better advertising copy? Although our estimates cannot directly answer these questions, our analysis and findings show that understanding AI’s broader impact will require researchers to investigate how different contexts and tasks vary in both the quality of AI’s output and the level of judgment people bring with them when using these tools.

2 Literature Review and Motivation

Recent experimental studies demonstrate that generative AI can effectively automate and enhance a wide range of well-defined knowledge work tasks. Noy and Zhang (2023) found that US white-collar workers who received generative AI training completed writing tasks approximately 40% faster while simultaneously improving output quality by 18%. Similarly, Dell’Acqua et al. (2023) studied BCG consultants worldwide and showed that those with GPT-4 assistance developed strategic plans 25% faster and received 40% higher quality ratings—with particularly strong benefits for consultants who previously struggled with such tasks. The benefits extend to creative work as well: Doshi and Hauser (2023) demonstrated that UK participants wrote more creative and engaging short stories when working with AI-generated ideas, while Meincke et al. (2024) found that consumer product ideas generated entirely by AI outperformed those created by students at an elite US university.¹ Collec-

¹Importantly, both studies noted that AI assistance led to more homogeneous outputs.

tively, these findings provide compelling evidence for AI’s substantial economic potential in knowledge work.

However, these promising task-level findings raise crucial questions about how AI will operate in real-world organizational settings. Unlike the precisely controlled experimental conditions in many studies, as Noy and Zhang (2023) observe, “real-world tasks involve vaguer objectives and instructions, requiring workers to exercise initiative in determining” where to apply AI tools. While some workplaces mirror these structured experimental settings, such as the standardized call center studied by Brynjolfsson, Li, and Raymond (2023), most organizations operate with far less structure. In these varied environments, from small businesses in India to AI R&D labs in San Francisco, workers must first decide whether and where to deploy AI assistance. This discretion matters significantly: Dell’Acqua et al. (2023) document that AI’s impact can be strongly positive or negative depending on the specific tasks for which it is employed. Consequently, variation in how much discretion workers and firms exercise over AI deployment may substantially shape both the technology’s estimated impact and its distributional consequences.

Consistent with these ideas, a substantial literature dating back to decision support systems research demonstrates that differences in judgment, incentives, and organizational context can undermine the benefits of machine-generated assistance. In many ways, today’s enthusiasm for AI mirrors the optimism of the 1970s and 1980s for decision support systems, when advocates argued that computer-based real-time interactive dialogue would yield superior decisions and outcomes (Keen and Scott Morton, 1978; Sprague Jr, 1980). Yet subsequent experiments and meta-analyses documented mixed results at best, with often limited improvements from decision support systems in real organizations (Sharda, Barr, and McDonnell, 1988). These disappointing outcomes frequently stemmed from the varied ways users incorporated tool outputs into their daily decision-making processes. Similarly, recent experiments examining AI integration show that relying on AI predictions can sometimes actually hinder performance (Kim et al., 2024; Agarwal et al., 2023). When workers lack the judgment to selectively implement the most useful aspects of AI recommendations, machine assistance may produce null or even counterproductive effects.

As a result, to understand AI’s potential impact on firms and workers, we must study

both well-defined knowledge tasks and “messier settings” where workers and firms exercise discretion over AI deployment and learning. Entrepreneurship offers an especially promising context for this broader evaluation. As Lazear (2004) argues, entrepreneurs perform an unusually wide variety of tasks, making them ideal subjects for understanding how AI’s impact might generalize beyond controlled experimental settings. Entrepreneurs must decide whether to use AI and for which specific tasks, then exercise judgment to effectively incorporate AI outputs into their businesses (Agrawal, Gans, and Goldfarb, 2022). Furthermore, Noy and Zhang (2023) note that when AI is applied to more open-ended, real-world “tasks and work requiring context-specific knowledge,” the distributional consequences may shift substantially, with AI potentially complementing and amplifying existing talent differences rather than substituting for expertise. These dynamics make entrepreneurship—including small-scale micro-businesses—particularly valuable for testing how AI’s distributional effects generalize, since business creation and management demand both contextual knowledge and open-ended decision making (Khanna, 2014).

Beyond understanding how AI’s impact generalizes to the important task of entrepreneurial decision making, simply testing AI’s impact in an emerging market is valuable in and of itself. Scholars have speculated that limited training data from emerging markets may cause LLMs to provide less useful or erroneous advice in such settings (Björkegren, 2023; Tao et al., 2023). Even when relevant data exist, fundamental differences between emerging and developed economies—in the tasks firms typically perform, worker human capital, and decision-making judgment—may cause AI to have substantially different impacts on business performance. The skill frontier dynamics observed by Doshi and Hauser (2023) illustrate this challenge: in their experiment, high performers gained little from AI because they already operated at the knowledge frontier. This finding suggests that in developing economies, where most workers are far from the skill frontier, AI adoption patterns and beneficiaries may differ dramatically from those observed in developed markets. Given that an estimated 85% of the world’s population resides outside developed economies, understanding how AI performs in these settings is crucial for assessing the technology’s broader economic potential. The policy implications are particularly significant: if AI can effectively substitute for scarce business expertise in resource-constrained developing economies, this technology may

offer unprecedented opportunities to upskill entrepreneurs and workers in regions where such interventions could yield the greatest returns to economic development (McKenzie, 2021).

3 Experimental Design and Analysis

This section outlines our experimental design and analytic strategy. We first provide background on the study context and recruitment process. We then review the design of the two experimental interventions and our main business performance outcome. Finally, we present regression equations used to estimate the impact of the generative AI assistant on firm performance.

3.1 Study Setting

Our study was carried out in the East African country of Kenya, where about 80% of employment takes place in the informal sector (Kenya National Bureau of Statistics, 2023). Kenya is also a leading adopter of digital technologies in Africa. For instance, it was one of the first countries to have large-scale adoption of mobile money, with M-Pesa—launched in 2007—reaching more than 34 million average monthly users in 2024 (Safaricom PLC, 2024). Kenya also has had one of the highest cell phone ownership rates in the region, with approximately 95% of adults owning a mobile phone (Kamau et al., 2024), most of whom use WhatsApp for text communication (Shengale, 2025).

Kenya's widespread digital infrastructure and adoption make it an ideal setting for studying the impacts of a WhatsApp-based generative AI assistant, as participants are already familiar with interacting through such platforms. Additionally, without the foundation of widespread mobile access, the delivery of our intervention would not have been possible, highlighting an important boundary condition for applying similar interventions in other emerging market contexts (Callen et al., 2025).

3.2 Recruitment and Sample of Kenyan SMB Entrepreneurs

Our experiment was carried out in three stages. In the first stage, we recruited participants and administered three pre-treatment surveys. In the second stage, participants were randomly assigned to the generative AI business assistant group or a control group. In the final stage, participants were tracked over an additional four post-treatment survey waves. The full experiment timeline is shown in Figure 1.

We began recruiting for the pre-treatment surveys on 21 May 2023 through ads run on the Meta ad platform in partnership with the Busara Center, a Kenyan research organization with more than a decade of experience. These ads invited entrepreneurs to complete a short survey. The purpose of this initial survey, along with two subsequent pre-treatment surveys, was to identify and filter out respondents who did not own and operate a business (i.e., who were not SMB entrepreneurs), who failed to complete all three surveys and were therefore likely to attrit over time, or who did not correctly answer basic attention check questions. Participants received a small amount of compensation for completing each survey, ranging from 30 Kenyan Shillings for the initial screening survey to 200 Kenyan Shillings in each of the four post-treatment surveys.

This recruitment and screening process produced a final sample of 640 Kenya SMB entrepreneurs. The sample is geographically diverse: participants came from 44 of Kenya's 47 counties, with the largest shares from Nairobi (133), Kiambu (53), and Nakuru (47) (see Figure A1). The median entrepreneur was 26 years old, and most had at least some college education, with only 10% reporting no college experience. Women made up 32% of the sample, consistent with lower rates of female business ownership in Kenya ([The World Bank Group, 2018](#)). The businesses represented in the sample ranged from fast food vendors to poultry farmers to cybershop owners (who provide services such as internet access, printing, scanning, and computer support). The most common sectors in the sample were Food and Beverage (27%), Clothing (21%), and Agriculture (19%).

The median entrepreneur in the sample had been running their business for one year or less and had no employees. Reflecting both the emerging-market context and these firms' limited operating histories, entrepreneurs reported modest pre-treatment business perfor-

mance. The average pre-treatment monthly profit was about KSh 19,550, and reported monthly profits ranged from about KSh 2,650 at the 5th percentile to about KSh 55,100 at the 95th percentile.

3.3 Experimental Interventions

Following the third pre-treatment survey, participants were randomly assigned to either the treatment group—which received access to the generative AI-powered business assistant—or the control group, which received access to a set of publicly available business training guides, both of which are discussed in more detail below. To improve the statistical precision of our experiment, we stratified random assignment by quartiles of pre-treatment business performance and by gender. As expected under randomization, the two groups were comparable across a range of baseline characteristics and performance measures ([Table A1](#)).²

Before being given access to the AI assistant, participants in the AI condition completed a short (5–10 minute) online survey-based training. This training explained that the chatbot they would be interacting with was an AI system (not a human), that the AI’s information would not be up to date, and that the advice it provided could sometimes be inaccurate. To ensure comparability across conditions, a similar training was given to the control group—though this version focused on the business training guides and did not mention the AI tool. We provide details on the training materials for both groups in [Subsection C.2](#).

3.3.1 Treatment (The AI Business Assistant)

The generative AI-powered business assistant was built on GPT-4, a large language model released by OpenAI in March 2023 ([OpenAI, 2023](#)). To ensure accessibility, the AI assistant was deployed through WhatsApp, a messaging platform already widely familiar to participants in Kenya. Access was restricted so that the assistant would only respond to phone numbers belonging to participants in the treatment group, ensuring that participants in the

²post-treatment attrition was also similar across conditions. We classify participants as having attrited if they failed to complete the final survey, even if they completed some earlier post-treatment surveys. Of the 640 study participants, 11 attrited from the AI-assistant group and 7 from the control group. The 1.2 percentage point difference in response rates between the two conditions is not statistically significant ($p = 0.35$).

control group could not use the tool.

Interacting with the assistant differed from directly using ChatGPT in three key ways. First, we implemented a custom *system prompt*—a set of initial instructions provided to the model—to tailor its behavior to the context of Kenyan SMBs. For example, if a user asked ChatGPT how to raise funding for a new business, it might suggest seeking venture capital—a strategy common for high-growth startups but one that is unrealistic for low-resource Kenyan SMBs. In contrast, our AI assistant would suggest more contextually-appropriate options, such as borrowing from family and friends, joining an informal investment club (“chama”), or exploring other community-based lending institutions. The complete AI system prompt is provided in [Appendix C](#).

Second, the assistant was instructed to provide three to five enumerated recommendations in response to each question. For example, an entrepreneur asking how to adapt to increased competition with their restaurant might receive suggestions such as (1) differentiating the menu, (2) improving customer service, (3) setting up a loyalty program, (4) partnering with local events and sponsors, and (5) maintaining high levels of cleanliness. Each recommendation from the AI assistant was accompanied by a short explanation. For instance, the loyalty program was described as a program “where customers earn points for every purchase which they can later redeem for free meals or discounts. It encourages repeat business as customers will want to accumulate more points.”

A key distinction between a generative AI assistant and a static business training course or guidebook is in how interactive they are: entrepreneurs could ask follow-up questions, explore alternatives, and learn more before deciding whether to implement advice. To support this, the third feature of the AI assistant allowed users to request additional details by entering the number that corresponded to a specific recommendation. Continuing the example above, an entrepreneur who entered the number 3 (for “loyalty program”) could receive more granular suggestions of different types of loyalty programs (e.g., “Point-Based System: Implement a simple point-based system where customers earn points for every shilling spent. Once they accumulate a certain number of points, they can redeem them for free meals or discounts. You could use simple tools like punch cards or digital apps that are inexpensive to manage.”). This feature of the AI assistant design encouraged deeper engagement and

follow-up exploration. An example interaction between a participant and the AI assistant is provided in [Figure 2](#).

Consistent with the many challenges facing the owners of young businesses, entrepreneurs used the AI assistant to discuss a wide range of topics. Conversations included how to motivate employees, allocate resources when expanding a store, adopt best practices in poultry farming, and manage bankruptcy. In [Appendix D](#), we present two detailed conversation logs: a restaurant owner (described above) navigating an increasingly competitive market, and a dairy salesperson coping with price hikes from their supplier. In both cases, entrepreneurs described their challenges to the AI assistant, considered different options, and returned for additional advice as new issues arose.

3.3.2 Control (Business Training Guides)

There are at least two challenges in measuring the causal impact of a generative AI-powered business assistant that shaped the design of the control group in our experiment. The first is conceptual: AI could often provide multiple benefits, making it difficult to define a precise counterfactual. For example, our AI assistant not only offers personalized advice but also encourages entrepreneurs to reflect on their businesses. As a result, any observed treatment effects could reflect the value of personalized support, the benefits of structured reflection, or both, making it challenging to isolate the specific contribution of the AI tool. We also wanted to send simple reminders to encourage use of the AI assistant. However, if only one group received reminders, estimated effects could conflate the impact of the reminders with the effect of the assistant itself ([Zwane et al., 2011](#)).

A second challenge is measurement. As in many studies conducted in emerging markets, we measure business performance through self-reported surveys. This introduces the risk of demand effects—changes in behavior or reporting driven by perceived expectations ([Zizzo, 2010](#)). If one group receives an intervention focused on their business while the other receives nothing, differences in survey responses may partly reflect expectations rather than actual performance change.

As a result, rather than using a “pure” (no intervention) control group, we designed the control condition to mirror the treatment group as closely as possible. Specifically,

participants assigned to the control group received access to a series of business training guides that were also sent over WhatsApp. These guides, developed by the International Labour Organization for SMB entrepreneurs in emerging markets, focused on strategies for business development and growth (ILO, 2014). Control group participants were sent links to three PDFs from a nine-part series, along with a link to access the remaining materials.³ Importantly, because existing evidence has found that delivering static training content over text messages has little effect on firm performance, we interpret this intervention as a placebo control condition (Cusolito, Darova, and McKenzie, 2023; Mehmood, 2023; Davies et al., 2024).

3.4 Outcomes

To measure firm performance, we follow approaches commonly used in emerging market contexts: collecting multiple self-reported performance measures in both pre-treatment and post-treatment survey waves (McKenzie, 2012), and constructing indices to reduce measurement noise (Kling, Liebman, and Katz, 2007). Specifically, in each of the three pre-treatment and four post-treatment waves, we collected standard self-reported measures of weekly and monthly profits and revenue (De Mel, McKenzie, and Woodruff, 2009). For example, when eliciting weekly [monthly] profits, respondents were asked:

Now we would like to ask you about the total profits across all of your businesses in the last 7 [30] days. What was the total income the business earned during the last 7 [30] days after paying all expenses including the wages of employees, but not including any income you paid yourself.

We asked a parallel set of questions for weekly and monthly revenue (the full set of questions is provided in Appendix E).⁴ We collected a total of 4,434 performance observations for each of the four business performance index components, of which 2,226 were from the

³For reference, all of these guides can be viewed using the following link: https://www.ilo.org/empent/areas/start-and-improve-your-business/WCMS_192062/lang--en/index.htm.

⁴These measures, along with a series of outcomes designed to examine experimental mechanisms, were prespecified following the third pre-treatment wave at <https://osf.io/64kws/files/osfstORAGE/64e386ae62e5ee196a3ab9c9>

AI-assistant condition and 2,208 were from the control group (summary statistics for the outcome variables are provided in [Table A2](#)).

Our primary outcome is a business performance index that aggregates weekly and monthly profits and revenues. The goal of this composite index is to reduce the noise that is common in self-reported performance data, and such indices are widely used in survey-based empirical research ([Kling, Liebman, and Katz, 2007](#)) and in emerging markets ([Campos et al., 2017; Anderson and McKenzie, 2022](#)). Additionally, aggregating multiple performance measures into a single index helps to mitigate concerns about multiple hypothesis testing. We constructed the standardized index following the procedure outlined in [Kling, Liebman, and Katz \(2007\)](#). First, we winsorized each of the four measures (at either the 99% or 95% level, depending on specification) to limit the influence of outliers. Second, we standardized each variable by subtracting the mean and dividing by the standard deviation of the control group’s post-treatment outcomes, which puts all four index components on a common standardized scale. Finally, we averaged the four standardized components in each survey wave to create the index.

Comparing the average values of the performance index among treatment and control participants provides an aggregate measure of how the AI system impacts firm performance, measured in standard deviations. If the difference in the performance index between the AI assistant condition and the control group is greater than zero, it means that the AI system led to improvements in firm performance relative to control; if it is less than zero, it indicates a decline in business performance. The magnitude of these effects can be interpreted as standard deviations of performance improvement: for example, a treatment effect of 0.2 indicates that the AI system improved business performance by 0.2 standard deviations relative to the post-treatment control participants’ performance distribution.

3.5 Estimation Strategy

We estimate the average treatment effect of the generative AI assistant on the business performance index using an OLS regression:

$$y_{i,t} = \beta_0 + \beta_1 \mathbb{1}(\text{AI}_i) + \delta y_{i,-t} + \mathbf{X}_i + \tau_t + \theta_i + \varepsilon_{i,t} \quad (1)$$

where $y_{i,t}$ is the business performance index for participant i in post-treatment period t . $\mathbb{1}(\text{AI}_i) \in \{0, 1\}$ is an indicator equal to 1 if participant i was assigned to the AI assistant, and the coefficient β_1 captures the average treatment effect of the AI assistant relative to the control group. This regression also controls for pre-treatment values of the outcome variable ($y_{i,-t}$), which can increase statistical power in randomized experiments compared to a difference-in-difference approach (McKenzie, 2012). In addition, the specification includes time-period fixed-effects τ_t , strata fixed effects θ_i , and a broader set of controls \mathbf{X}_i selected using the double-LASSO procedure of Belloni, Chernozhukov, and Hansen (2014). Candidate covariates for selection using the LASSO procedure included pre-treatment management practices, oldest business age, sector, count of children, age, education, gender, ChatGPT use, personality measures (agreeableness, conscientiousness, extroversion, openness, neuroticism), and Kenyan county.

To estimate heterogeneous treatment effects on our standardized business performance index, we use the following equation:

$$y_{i,t} = \beta_0 + \beta_1 \mathbb{1}(\text{AI}_i) + \beta_2 \text{Het}_i + \beta_3 (\mathbb{1}(\text{AI}_i) \times \text{Het}_i) + \delta y_{i,-t} + \mathbf{X}_i + \tau_t + \theta_i + \varepsilon_{i,t} \quad (2)$$

where Het_i represents a prespecified dimension of heterogeneity, and β_3 captures the differential effect of AI assistant access along that dimension. We examine heterogeneous treatment effects with respect to each of the following three variables: (i) baseline business performance, (ii) gender, and (iii) prior ChatGPT use. Baseline business performance is defined as a median split of the standardized performance index across the three pre-treatment periods, and frequent ChatGPT use is defined as reporting use of ChatGPT at least once per week prior to the experiment.

4 Results

4.1 Impacts of the AI Assistant on Firm Performance

Our first set of results examines the impact of access to the generative AI assistant on the standardized business performance index. We find no statistically significant average treatment effect of access to the AI assistant on firm performance (winsorized at the 99% level) relative to control, as shown in Estimate A in [Figure 3](#) and [Table 1](#) ($\beta = 0.04$ standard deviations (sd), $p = 0.34$, 95% CI = [-0.05, 0.15]). In other words, in the context of small and medium-sized Kenyan businesses, we do not find any evidence that providing access to a generative AI assistant improves firm performance *on average*.

4.2 Treatment Effect Heterogeneity

Building on recent studies showing that AI assistance especially helps lower performing workers ([Dell'Acqua et al., 2023](#); [Noy and Zhang, 2023](#); [Peng et al., 2023](#); [Brynjolfsson, Li, and Raymond, 2025](#)), we next estimate heterogeneous treatment effects by splitting our sample into two groups, based on the pre-treatment values of the same performance index we use as our primary outcome variable.⁵ We refer to entrepreneurs who had below-median pre-treatment performance as *low performers* and those who had above median pre-treatment performance as *high performers*.

Estimates B and C in [Figure 3](#) and [Table 1](#) show the estimated treatment effect within the low- and high-performer sub-samples. For low performers, access to the AI assistant reduced average business performance relative to the control group ($\beta = -0.08$ s.d., $p = 0.01$, 95% CI = [-0.13, -0.03]). Because our index is standardized against the standard deviation of the post-treatment control group ([Kling, Liebman, and Katz, 2007](#)), we can contextualize the magnitude of our effect by looking at the mean and standard deviation of revenues and profits for this group. For example, [Table A2](#) shows average monthly revenues are 51,392 Kenyan Shillings (USD \$360.90) with a standard deviation of 59,353 Shillings.⁶ Multiplying

⁵We also prespecified testing for heterogeneity with respect to gender and prior use of ChatGPT. We report the results of these tests in [Figure A2](#). We find no evidence of treatment effect heterogeneity with respect to either of these dimensions.

⁶We convert to USD using the exchange rate at roughly the midway point of our experiment, July 31st,

the standard deviation by our -0.08 s.d. estimate suggests that the AI assistant caused a drop of 4,748 Shillings (\$33.34) in revenue per month; a 9.2% relative decrease. Regardless of whether we perform this contextualization exercise with pre-treatment measures or post-treatment measures, revenue or profits, they consistently yield a negative treatment effect estimate of just under 10% for low performers.

For high performers, we find suggestive evidence that AI assistance improved performance relative to control ($\beta = 0.16$ s.d., $p = 0.07$, 95% CI = $[-0.02, 0.33]$). Again we multiply the standard deviation of monthly revenue by our estimate, now a 0.16 s.d. increase, which suggests that AI caused an increase of 9,496 Shillings (USD \$66.69) in monthly revenue; a 18% relative increase. Again, whether we use pre-treatment and post-treatment measures to contextualize the effects' sizes, they consistently produce a positive treatment effect estimate of over 15% for high performers.

Building on these subgroup analyses, Estimate D in [Figure 3](#) and [Table 1](#) directly compares the impact of the AI assistant on initially low and high performers using the heterogeneous treatment effects model in [Equation 2](#). We find strong evidence that the effect of AI is different for low and high performers. Initially, low-performing entrepreneurs experienced a 0.23 sd lower causal effect ($p = 0.01$, 95% CI = $[-0.41, -0.05]$) of the AI assistant compared to high performers.⁷ In monetary terms, the estimated difference in effect size between the two groups represents a change in in the impact of AI on a firm's monthly revenue of roughly 13,651 Shillings (USD \$95.86).

4.3 Robustness

Next, we test the robustness of our findings. We begin by examining whether the observed differences in treatment effects between initially low and high performers are driven by a few outliers or reflect broader shifts in the performance distributions. [Figure 4](#) plots the residualized distribution of business performance among initial low performing firms (Panel A) and initially high performing firms (Panel B). In both cases, there are noticeable differ-

²⁰²³, using [Wise's KES to USD historical rate calculator](#).

⁷Given this uneven effect, and especially the negative performance effect we find for low performers, in [Appendix F](#) we discuss ethical considerations and the importance of policy equitope in designing and interpreting experiments.

ences between the treated and control empirical cumulative distribution functions (eCDFs): in Panel A, the control eCDF is shifted to the right of the treated participants' eCDF, indicating higher performance, while in Panel B, this pattern is reversed. These results indicate that the observed differences in treatment effects are not driven by a few extreme cases, but rather reflect broader shifts in the distributions of profits and revenues for low- and high-performing firms.

Having established that our heterogeneous effects reflect genuine distributional shifts, we next examine whether our results hold across alternative specifications. [Figure A3](#) shows that the observed pattern of results—a null average effect, which masks significant heterogeneity by initial business performance—holds when we winsorize our business performance outcome at the 95% instead of the 99% level and when we exclude participants who did not complete all post-treatment surveys. Across these specifications, our most conservative estimate of treatment effect heterogeneity is 0.18 s.d. ($p < 0.05$, 95% CI = [0.02, 0.34]).⁸ Figures [A5](#) and [A6](#) show that we find qualitatively similar results when looking at individual components of our business performance index. We also show in Appendix [G](#) that our effects are neither amplified nor muted by spillovers between participants, and in Appendix [A7](#) we plot survey-wave specific performance estimates. Consistent with our main results, performance only starts to diverge for both low- and high-performers post-treatment.

Finally, we apply multiple testing corrections to our results. In doing so, we show that our estimates are not simply explained by the fact that we have tested for performance treatment effects multiple times, increasing the chance that any given statistically significant estimate is a false positive. We use the Benjamini-Hochberg procedure ([Benjamini and Hochberg, 1995](#)) to control the false discovery rate—the expected proportion of false positives among all results deemed statistically significant—across three key families of results: *(i)* the main performance effect estimates, *(ii)* the heterogeneous treatment effect estimates, and *(iii)* the effects on our survey-based mechanism measures. We report p -values adjusted for multiple comparisons in [Table A3](#), while noting that our key results remain statistically significant

⁸Our weakest performance results come not from different levels of winsorization or adjustments to our sample, but from applying an inverse hyperbolic sine function to our performance index. The results of these analyses are shown in [Figure A4](#). In Appendix [E.2](#) we discuss why our performance measure is not appropriate for this transformation ([Chen and Roth, 2024](#)).

even after adjusting for multiple comparisons.

5 Unpacking Mechanisms

To better understand why low and high performers exhibit different performance effects, we follow an abductive approach and use the rich text data we collected from interactions with the AI assistant and our surveys to shed light on the potential drivers of our starkly uneven performance effects (Pillai, Goldfarb, and Kirsch, 2024). As noted in our literature review, prior research on generative AI’s economic impacts typically involves structured tasks. In contrast, our participants had considerable discretion in how and when they used AI assistance. Participants not only had access to AI output, but could choose to apply it to myriad different tasks. They then had to exercise considerable judgment in deciding when to implement the AI’s advice. To understand whether and how this discretion shaped the impact of AI, we structure our analysis of the mechanisms into three parts. First, we look at the questions entrepreneurs asked, testing for differences in the number and types of questions asked by low- and high-performing entrepreneurs. Second, we turn to the AI itself, testing if the AI’s responses differed across groups. Third, we examine whether entrepreneurs implemented the AI’s advice and, if so, whether low and high performers chose different types of suggestions to implement.

5.1 Data and Analytical Methods

We use three different data sources to explore the mechanisms driving our heterogeneous performance effects.⁹ First, to test for differences in the quantity and types of questions asked, we analyze the complete set of 4,810 messages sent by the 275 treated entrepreneurs who engaged with the AI assistant at least once (47 entrepreneurs simply never asked the AI a question). Almost 90% of the entrepreneurs sent the AI at least one message with the

⁹We also prespecified alternative dependent variables based on post-treatment multiple-choice survey questions to test a host of potential mechanisms. This analysis, described in Subsection E.3, examines differences in areas such as management practices and the level of stigma entrepreneurs feel when seeking advice. The results of these analyses are found in Figure A8. Across the seven mechanisms tested, we do not find any evidence of statistically significant heterogeneous treatment effects. This absence of clear heterogeneity in our prespecified mechanism measures motivates much of the text analysis.

average number sent being just over 17 messages. Given that many messages are simple numeric follow-ups or fragments (e.g. pressing “3” to learn more about option 3 or a typed reply of “y”), we focus our text analysis on 1,392 substantive business messages that exclude numeric follow-ups, fragments, and non-business content as determined by an independent pair of human coders (see Appendix [Subsection H.1](#) for a more detailed description of our text data).

Second, to explore differences in how the AI responds, we turn to corresponding data on the AI assistant’s responses. As depicted in [Figure 2](#), our AI assistant typically provided three to five recommendations per response, with each entrepreneur receiving an average of nearly 24 such suggestions. [Table A4](#) presents the 100 most frequently occurring recommendation categories from a total of 4,207 distinct suggestions generated by the AI assistant. These categories are diverse and span topics including: “Community Engagement” (59 occurrences, the most frequent), “Location” (28), “Diversify Your Product Range” (25), “Social Media Marketing” (17), “Promotions and Discounts” (8), and “Employee Training” (6).

Third, to test how entrepreneurs used AI advice, we use text responses from the four post-treatment surveys that asked them to describe (1) their products, (2) services, (3) their processes, and (4) the most impactful business changes they had made over the past 30 days. We combine these four survey responses into a single text object in our analysis. Unlike the question and answer data, we have business changes text for both treated and control participants. [Table 2](#) shows that among treated participants, performance status does not predict whether they have available question or answer text, and among all participants, neither treatment status nor performance predicts whether they have business change text, allowing us to analyze differences in text content without selection concerns.

While the mechanisms we aim to test—Do entrepreneurs ask different types of questions? Are they making different types of changes to their business? Do the business changes depend on the AI’s advice?—are conceptually straightforward, translating these ideas into quantitative tests is nontrivial for two reasons. First, a piece of text does not vary along a single dimension (e.g. is it about farming?) but along a myriad of semantic dimensions. Thus, understanding if low- and high-performers ask different types of question requires generating a multitude of quantitative semantic measures from the text and then testing if there

are any meaningful differences across all of these variables. Second, testing if entrepreneurs follow the AI’s advice requires testing if the content of the AI generated text *causes* changes in the content of the entrepreneur’s reported business changes. Yet, the AI’s advice is far from exogenous; by construction it depends on the very endogenous questions asked by the entrepreneur.

To address both concerns, we turn to recent advances in statistics and machine learning. To generate quantitative measures of the text’s content, we use text embeddings to capture the semantic meaning of text in high-dimensional space. We then use flexible machine learning prediction models to test if these vectors can predict whether an entrepreneur is a low or high performer, allowing us to run a single test to see whether there are any semantic differences in text produced by the two groups. To isolate the causal impact of AI’s advice on business changes, we couple word embedding methods with Pearl’s backdoor criterion to test for causal relationships in semantic space. In the following subsections, we provide brief explanations of each of these methods before turning to the results.

5.1.1 Text Embeddings

Text embeddings are vector representations of text that capture semantic meaning in high-dimensional space. Unlike simple word counts or bag-of-words approaches, embeddings encode relationships between concepts such that semantically similar texts (e.g., “reduce prices” and “offer discounts”) have similar vector representations even when they share no common words ([Mikolov et al., 2013](#); [Pennington, Socher, and Manning, 2014](#)).

The embeddings we use are generated by neural language models trained on massive text corpora through self-supervised learning ([Devlin et al., 2018](#); [Radford et al., 2019](#)). These models learn to predict missing or subsequent words in text, leading them to develop internal representations that capture semantic and syntactic relationships. The resulting embeddings encode information about word meanings, grammatical structures, and contextual usage patterns. Here, we generate 3,072-dimensional embedding vectors for all text using OpenAI’s `text-embedding-3-large` model, a high-quality embedding model that provides robust semantic representations ([OpenAI, 2024](#)). Each dimension in these vectors corresponds to learned patterns from large-scale text data, capturing abstract linguistic

features, though individual dimensions lack straightforward interpretation. Our approach extends recent management research that has used text embeddings to represent complex semantic relationships while maintaining computational tractability for downstream analysis (Cao, Koning, and Nanda, 2023; Goldberg and Srivastava, 2022).

5.1.2 Machine Learning Prediction Models

Having generated 3,072-dimensional word embedding vectors that represent the text from high- and low-performers, how can we parsimoniously test for differences across different groups? We employ machine learning prediction models to test whether the word embeddings representing our data can be used to distinguish between entrepreneurs who are high- or low-performers or from different treatment conditions. This approach treats text classification as a supervised learning problem, in which we attempt to predict known entrepreneur characteristics based on underlying text features. For example, if low performers ask about “two for one promotion,” “lowering prices,” and other “discounting” related ideas more often, then the embedding element(s) that map to discounting will on average be larger for this group, enabling the machine learning model to more accurately predict if an entrepreneur is a low or high performer. As a consequence, when a model’s prediction accuracy is significantly above random chance, this indicates that there are meaningful, systematic differences that exist in the underlying text between the groups. When the model cannot make accurate predictions, this indicates there are no meaningful differences between the groups across any of the multitude of semantic characteristics captured by the word embedding vectors used in our analysis.

Specifically, we use random forest models (Breiman, 2001), which make predictions by averaging results from many individual decision trees. By incorporating dual randomization—both across observations and features included in each tree—this approach reduces overfitting while maintaining predictive power. Random forests are particularly well suited for high-dimensional settings such as ours, where features outnumber observations, because they are able to handle many features without requiring a priori feature selection, can capture complex non-linear interactions, and provide robust predictions even when many features are irrelevant or noisy (Hastie, Tibshirani, and Friedman, 2009). Given our relatively small

sample size, we use 60% of our data for training and hold the remaining 40% as our test set. We implement hyperparameter tuning using the caret package in R with nested validation. Following Kuhn (2008), we fit models with 500 trees and a minimum node size of 1, then use 5-fold cross-validation within the training set to determine the optimal number of randomly sampled predictors.¹⁰

5.1.3 Causal Identification via Pearl’s Backdoor Criterion

To estimate the causal impact of the AI assistant’s advice on the types of business changes made by entrepreneurs in our sample, we apply Pearl’s backdoor criterion (Pearl, 2009), a causal identification strategy that uses directed acyclic graphs (DAGs) to isolate treatment effects by controlling for confounding variables. The backdoor criterion provides conditions under which observational data can be used to identify causal effects without relying on strong parametric assumptions.

In our context, we face a potential confounding problem: entrepreneurs’ underlying characteristics—such as their background, skills, industry context, and business acumen—could independently influence both the AI’s advice and the entrepreneur’s subsequent business decisions, creating spurious correlation between the text of the AI advice and the text describing the types of business changes. For example, an entrepreneur who is already launching a new marketing campaign could ask the AI about “advertising strategies,” receive feedback on advertising from the AI, and then note that they had launched an ad campaign. As this example illustrates, to estimate the causal impact of the AI on the types of business changes an entrepreneur makes, it is not enough to show there is a correlation between the content of the AI’s advice and the business changes described by the entrepreneur.

The backdoor criterion states that we *can* identify the causal effect of treatment X (AI advice content) on outcome Y (business changes), if we can control for a set of variables Z that: (1) blocks all “backdoor paths” from X to Y , i.e., all paths that connect X and Y but do not go through the causal pathway of interest, and (2) does not include any variables that

¹⁰We test models with different numbers of randomly sampled predictors set to 2, $\sqrt{p}/2$, and \sqrt{p} , where p is the total number of predictors (3,072 in our case), selecting the hyperparameter value that maximizes cross-validation accuracy within the training set (Kuhn, 2008). The final model with optimal hyperparameters is then evaluated on the held-out test set.

are themselves affected by X . In the vast majority of settings, Z is not fully measured, which undermines this identification strategy. However, in our context, because the AI assistant has no information about entrepreneurs beyond what they communicate in their messages, entrepreneur characteristics can only influence AI advice through the questions asked. As we show in [Figure 5](#), by controlling for the content of the questions (our Z), we can fully close the backdoor path through which the characteristics of entrepreneurs could confound the causal relationship between AI advice and business changes.

Specifically, we turn to word embeddings to quantitatively represent the content of the text data. We run a regression where each entrepreneur-embedding dimension pair is an observation, regressing the business changes embedding value for dimension d on the AI advice embedding value for that same dimension d , while controlling for the question embedding value for dimension d . This approach closes the backdoor path by ensuring that when we examine how AI advice content in any given semantic dimension affects business changes in that same dimension, we account for the entrepreneur’s initial question content in that dimension.

5.2 Testing for Differences in Question Asking

We first test whether low- and high-performing entrepreneurs differ in their likelihood or manner of interacting with the AI assistant. This analysis helps us understand whether the heterogeneous treatment effects that we observe could stem from different usage patterns of the AI tool itself. Overall, we find no statistically significant evidence that the volume of interactions that low and high performers have with the AI assistant are different. Panel A of [Table A5](#) shows that the two groups have a similar likelihood of sending at least one message (88.2% for low vs. 84.2% for high performers), ask similar numbers of questions (18.6 vs. 15.5 on average), pose a similar number of numeric follow-ups (5.34 vs. 4.57), and write questions of similar length (13.5 vs. 14.4 words).¹¹

To assess whether the *content* of the questions asked by low and high performers differs, we convert each entrepreneur’s entire set of questions into a 3,072-dimensional embedding

¹¹Median message counts are also nearly identical (9 vs. 8.5 messages for low and high performers, respectively).

vector and conduct analysis using these embeddings. As a reminder, in embedding space, similar questions (such as *How can I effectively market my vegetable business online?* and *What social media strategies will attract more customers to my vegetable business?*) are located close to each other, while semantically different questions (like *How do I treat pests in my maize crop?* and *How do I manage employee payroll records?*) are represented by vectors that are farther apart. If high- and low-performing entrepreneurs differed meaningfully in the questions they asked, their question embedding vectors should be predictive of their pre-treatment business performance. We test this by training a random forest model using the 3,072-dimensional embedding vector to predict whether an entrepreneur was a low- or high-performer. Table 3 shows that our model achieves an accuracy of only 43.2% on the test set ($p = 0.97$, 95% CI = [0.34, 0.53]), which is not statistically distinguishable from the “no information” random-guess benchmark of 51.4% accuracy. In other words, we do not find any evidence that there are substantial differences between the questions asked by the two groups.¹²

5.3 Testing for Differences in AI advice

Although high and low performers ask similar questions, the AI’s responses might still differ systematically between these groups. Even when questions address similar topics, subtle differences in word choice, phrasing, or context could influence how the AI generates advice. Additionally, the AI might respond differently to questions of varying specificity or complexity. If such differences exist, they could explain the heterogeneous treatment effects we observe, as one group might systematically receive higher-quality or more actionable advice than the other.

We first test for differences in the *amount* of advice that high and low performers receive from the AI assistant. Panel B of Table A5 shows that low and high performers receive an average of 21.5 and 25.6 pieces of advice, respectively. The median number of suggestions is also similar: 13 for low performers and 11 for high performers. Additionally, the average

¹²Consistent with these findings, in Appendix Subsection H.2 we use large language models to develop a ten-topic taxonomy and classify entrepreneur questions into these topics. We find that low and high performers ask about similar topics.

length of AI responses is nearly identical (168 words for low performers, 166 words for high performers). Again, none of these differences are statistically significant.

To assess whether the *content* of the AI assistant’s responses differed, we trained a random forest to predict entrepreneurs’ performance status using embedding vectors representing the full set of AI-generated answers delivered to each entrepreneur. The model achieves an accuracy of only 51.8% (Table 3; $p = 0.54$, 95% CI = [0.42, 0.61]), which is exactly the same as the “no information” random-guess benchmark of 51.8%. Overall, there is no evidence for differences in the types of advice delivered to the two groups.¹³

5.4 Do Entrepreneurs Follow AI-Generated Suggestions?

Our analyses thus far have established that high and low performers interact with the AI assistant in remarkably similar ways: they engage at similar rates, ask comparable questions, and receive similar advice. Given that we are unable to detect statistically significant differences in the AI interactions of entrepreneurs, we next turn to the possibility that our heterogeneous treatment effects are driven by differences in the way entrepreneurs act on AI advice.

First, we test whether entrepreneurs, irrespective of whether they are low or high performers, follow the AI’s advice. Using the backdoor criterion identification strategy described in Subsubsection 5.1.3, Table 4 presents the results of regressions where the unit of observation is an entrepreneur-embedding dimension pair. We regress the value of each embedding dimension from the “business changes” text on the corresponding embedding dimension from the “AI Advice” text, clustering standard errors by entrepreneur and embedding dimension to account for dependencies within both.

Model 1 shows that a one standard deviation increase in a given AI advice embedding element correlates with a 0.26 standard deviation increase in the corresponding business changes embedding element ($p < 0.0001$). However, this correlation could reflect, e.g., entrepreneurs asking about topics they were already planning to address. Model 2 uses

¹³Consistent with these findings, in Appendix Subsection H.2 we use large language models to develop a twelve-topic taxonomy and classify AI assistant responses into these topics. We find that responses to low- and high-performers address similar topics.

the backdoor criterion to address this concern by controlling for the embedding value from the questions text. With this control, a one standard deviation increase in an AI advice embedding dimension leads to a 0.127 standard deviation increase in the same dimension for subsequent business changes ($p < 0.0001$, 95% CI = [0.1073, 0.1475]). The difference between Models 1 and 2 suggests that roughly half the correlation stems from confounders such as entrepreneurs' pre-existing plans and half from the AI's influence. Model 3 confirms robustness to entrepreneur fixed effects.¹⁴

We find no evidence that low and high performers differ in their likelihood of following AI advice. Model 4 interacts both answer and question embeddings with a high-performer indicator; neither interaction is statistically significant. Models 5 and 6 split the sample by pre-experiment business performance and show virtually identical estimates for both groups.

5.5 Do High and Low Performers Implement Different Types of Changes?

Given that questions asked of and answers provided by the AI do not systematically differ, but that both low and high performers act on the AI's advice, the remaining potential explanation for our heterogeneous treatment effects is differences in how entrepreneurs select and implement the advice they receive. To test for these differences, we turn to self-reported data collected in our surveys on the "business changes" made by entrepreneurs in our study. If low and high performers differ in how they act on AI suggestions, this should lead to measurable differences in the business changes they report making.

5.5.1 Evidence of Differential Implementation Patterns

We first test whether the business changes made by treated entrepreneurs differ systematically from those made by their counterparts in the control group. If entrepreneurs are implementing AI advice in meaningfully different ways, we should be able to detect these differences in the content of their reported business changes. Following the same approach we used to analyze question and answer data, we train random forest models on embedding

¹⁴The results of a complementary word-level analysis that supports these conclusions, but relies on stronger methodological assumptions is available in [Table A6](#).

representations of business changes text to predict entrepreneur characteristics. In this case, since we have data on business changes from both treated and control groups, we use all observations for which we have business changes text.

The results provide strong evidence of differential implementation patterns. Model 3 in Table 3 predicts whether an entrepreneur is a low or high performer based solely on their business changes text, achieving 58.4% accuracy—significantly better than the 50.5% no-information benchmark ($p = 0.012$, 95% CI = [0.52, 0.65]). This suggests that low and high performers make systematically different types of changes to their businesses. Model 4 shows we can also predict treatment status with 59.8% accuracy ($p = 0.004$, 95% CI = [0.53, 0.66]), indicating that treated entrepreneurs make different changes than control entrepreneurs.

Most tellingly, Model 5 simultaneously predicts both performance level and treatment status (four categories total), achieving 33.8% accuracy compared to a 25.4% baseline ($p = 0.004$, 95% CI = [0.27, 0.41]). This suggests that the interaction between performance level and AI access creates distinct patterns of business changes—exactly what we would expect if high and low performers are implementing AI advice differently.¹⁵

5.5.2 Developing an Abductive Interpretation: Tailored vs. Generic Implementation

Though these results indicate that there are differences in the changes entrepreneurs implement, they do not tell us the content of these changes. To develop a theoretical understanding of these differences, we employ an abductive approach, examining the data to identify patterns and then developing interpretations that best explain these patterns (Pillai, Goldfarb, and Kirsch, 2024). To do so, we develop a novel empirical strategy to identify the *types* of changes that best represent the average differences between groups. The core idea is to find individual examples that best illustrate the systematic patterns we observe in the embedding space. We first calculate the average difference in business change embeddings between treatment and control groups separately for low and high performers, creating a "treatment effect embedding" for each group. We then identify individual pairs of control and treated

¹⁵The baseline accuracy is lower here because the empirical exercise involves predicting four classes rather than two.

business changes text whose embedding differences are most similar (using cosine similarity) to these average group differences. We select examples from the top 1% most similar pairs for illustration, essentially finding representative cases that best capture the systematic patterns while acknowledging we cannot observe individual counterfactuals (additional technical details in [Appendix H](#)).

Panel A of [Table 5](#) presents business changes descriptions that best represent the treatment effect for high performers. The left column shows examples from the control group, and the right from the treated group. Treated high performers often explicitly describe working with the AI assistant to implement specific changes: “getting advice from AI mentor on other ways of getting power when there is electricity power black out” or “renting out gaming accessories in my cybercafe following advice from your AI.” These examples vary considerably in content, consistent with [Figure A8](#), which shows no heterogeneous shifts across prespecified survey-based mechanism measures. Rather than applying generic advice, high-performing treated entrepreneurs appear to use the AI to make targeted, business-specific improvements ([Guzman and Li, 2023](#); [Carlson, 2023](#)). Panel B shows similar pairs for low performers. Treated low performers rarely mention the AI and more commonly report standard strategies such as discounting, lowering prices, or advertising. While these changes may reflect rational business decisions, they tend to be less differentiated and potentially less likely to lead to sustained improvements in firm performance ([Carlson, 2023](#)).

Through this exploratory and abductive analysis, our preferred interpretation emerges: high-performing entrepreneurs appear more likely to select and implement advice that is tailored and specific to their business context, while low-performing entrepreneurs tend to adopt more generic recommendations. We acknowledge this pattern could reflect alternative mechanisms, such as differences in implementation effort, willingness to undertake complex changes, or the time horizon over which different changes yield positive (or negative) results. However, the specificity and business-relevance of changes described by high performers, combined with their explicit mentions of AI guidance, points toward differential selection and implementation of contextually appropriate advice as a plausible explanation for our heterogeneous treatment effects.

5.5.3 Quantitative Evidence for the Tailored vs. Generic Hypothesis

To further test this interpretation, we develop quantitative measures of our proposed mechanism using keyword matching and word frequency analysis to measure AI-related engagement and specificity. We create three measures: First, we identify AI-related engagement by flagging text that explicitly references the AI assistant, chatbots, or artificial intelligence. Second, we assess specificity by counting words used by five or fewer entrepreneurs in our sample. These uncommon words (e.g., “150cc,” “arcade,” “biodegradable,” “peanut”) are more likely to reflect tailored, business-specific changes, in contrast to common terms like “advertising,” “customers,” or “price.” Third, we identify mentions of generic strategies, specifically discounting and advertising approaches.

Figure 6 summarizes our quantitative findings, which support the tailored vs. generic implementation hypothesis. Panel A shows that treated entrepreneurs are 6.3 percentage points more likely to reference working with or learning from the AI assistant ($p < 0.0001$, 95% CI = [3.336; 9.188]). This effect is concentrated among high performers, where the increase is 10.8 percentage points ($p < 0.001$), compared to a 1.8 point increase among low performers that is not statistically significant ($p = 0.19$), with the between-group difference statistically significant at $p < 0.01$. Panel B provides direct evidence for our tailored implementation hypothesis: treated high performers use 20% more uncommon words than those in the control group ($p < 0.05$, 95% CI = [1.04; 1.36]), consistent with more specific, tailored changes. Low performers, by contrast, use slightly fewer uncommon words, though the estimated treatment effect is quite small and not statistically significant. Panel C supports our generic implementation hypothesis: treated low performers are significantly more likely to report price cuts or advertising. Among low performers, 39% in the control group mention these strategies, compared to 51% in the treated group (a difference of 11.8 percentage points, $p < 0.05$, 95% CI = [0.8775; 22.78]). There is no statistically significant difference in the rate at which high performers mention price cuts or advertising. A more detailed description of these analyses can be found in Appendix Subsection H.5.

5.6 Generalizing from AI’s Uneven Impacts: The Role of Judgment and Ability

Beyond explaining the uneven performance impacts in our study, this mechanism analysis also points to a more general framework for understanding when AI has more or less uneven impacts. As noted in the literature review, our study differs from prior experiments testing the impact of AI on well-defined tasks in that entrepreneurs must choose for which of many open-ended tasks to use AI and then select which AI suggestions to implement. The results from our text analysis show that variation in which ideas entrepreneurs implement likely drives at least some of the uneven impact we observe. Low and high performers receive the same types of AI-generated suggestions, with both groups receiving a mix of generic and tailored ideas of varying quality. However, high performers are better at selecting and implementing the more promising, tailored suggestions. As illustrated in Figure 7, when an AI suggests both good and bad ideas, this opens the possibility that differences in an entrepreneur’s judgment and their firm’s capabilities will matter (Agrawal, Gans, and Stern, 2021; Agrawal, Gans, and Goldfarb, 2022). In this stylized diagram, an entrepreneur who blindly follows AI advice will on average perform worse, while an entrepreneur who knows which advice is promising and has the ability to implement it will perform better.

This illustration also provides a way to understand how AI’s impact might generalize. When AI is applied to a task that is well-defined and within the “jagged frontier,” the distribution of potential impact in Figure 7 shifts to the right with the output being uniformly helpful (Dell’Acqua et al., 2023). As a consequence, anyone can benefit from the AI’s suggestions even if they simply randomly draw from the distribution, and these benefits are likely to be largest for those who previously struggled to filter good from bad ideas or lacked the ability to produce high-quality output. When AI is applied to tasks where the distribution is shifted to the left, performance can drop across the board, as shown in Dell’Acqua et al. (2023). As in our study and consistent with Noy and Zhang (2023)’s concluding arguments, when AI is used for many tasks—and so is likely to generate a mix of good and bad suggestions—we should expect those with more judgment and ability to benefit more.

The stylized model in Figure 7 also helps highlight the promise and peril of AI for

workers and firms in emerging markets. Although experimental estimates of AI’s impact on performance are largely positive, these findings are from developed economies. In emerging markets, a lack of training data from such places may lead to a persistent shift to the left in the quality of AI-generated output, increasing the need for judgment in such markets (Björkegren, 2023; Tao et al., 2023). Compounding these challenges, workers and firms in emerging markets face human capital and capability gaps that may further dampen or even reverse the benefits of AI assistance. As these conjectures make clear, understanding variation in how people and businesses engage with AI—from what questions they ask to the types of AI suggestions they can and choose to implement—will be crucial to understanding AI’s broader impact and ensuring this technology benefits all.

6 Discussion

Our study is—to our knowledge—the first randomized controlled trial testing the impact of generative AI on people and firms in developing economies. Many entrepreneurs in developing countries face barriers to getting helpful personalized feedback, mentorship, and training (Chatterji et al., 2019; Dimitriadiis and Koning, 2024; Björkegren, 2023). Our findings suggest that generative AI-based tools can both harm and potentially help entrepreneurs in low- and middle-income countries, depending on their baseline performance. For high performers, we find suggestive evidence that access to generative AI may have driven improvements comparable in magnitude to effective business training programs, but delivered at a fraction of the cost.¹⁶ However, this potential benefit comes with an important caveat: we find robust evidence that generative AI can significantly harm low-performing businesses, with the AI assistant leading to nearly 10% worse performance for this group. More broadly, our findings suggest that while generative AI has the potential to benefit firms in developing countries (Björkegren, 2023; Lou, Sun, and Sun, 2023), untargeted deployment risks widening performance differences.

Previous studies have focused on AI assistance with specific, well-defined tasks where

¹⁶Similar to recent work highlighting the cost-effectiveness of text message interventions (Fabregas et al., 2024), our AI assistant costs a few dollars per participant, whereas most training programs cost hundreds of dollars per participant (McKenzie, 2021).

the distribution of advice quality is relatively narrow; either clearly beneficial or clearly harmful (Brynjolfsson, Li, and Raymond, 2025; Dell’Acqua et al., 2023). In contrast, our intervention allows entrepreneurs broad discretion in how they engage with AI, creating a much wider distribution of potential advice quality and implementation outcomes. Our mechanism analysis suggests that this broader distribution requires entrepreneurs to exercise judgment when selecting which advice to implement. High performers appear more capable of identifying and acting on contextually appropriate suggestions, while low performers systematically adopt strategies that reduce business performance: generic advice focused on price cuts and advertising that may be inappropriate for their specific business situations. This finding extends prior work by providing empirical evidence on how AI’s impact depends critically on user judgment and selection capabilities when the advice space is open-ended rather than constrained, with implications for the design and deployment of AI systems in complex, real-world environments.

6.1 Limitations

Our study has several important limitations. First, our findings come from a single context (small business entrepreneurs in Kenya), cover a limited time span (5 months), and test only a specific implementation of AI assistance (a carefully prompt engineered GPT-4 instance). In our context, even the highest-performing firms operate with relatively limited resources compared to businesses in developed economies. In developed economies, entrepreneurs often have greater access to education, expert advisors, and professional services, which may already provide a richer information environment than our AI assistant. Moreover, larger and more established firms are often less constrained by capital, enabling them to take greater risks and invest in more transformative AI-powered changes. This is not to suggest that our results lack evidential value—hundreds of millions of small businesses around the world resemble the microenterprises in our study. Rather, our findings speak to the need for a broader range of future empirical research that can unpack the global scope of generative AI’s impacts across firm size, sector, geography, access to capital, and other complementary investments that could enable firms to benefit from AI.

Turning to the study’s duration and timing, although the five-month duration of our

experiment is considerably longer than nearly all prior experiments measuring the economic impacts of generative AI (cf. Brynjolfsson, Li, and Raymond, 2025), there may still be longer-term impacts we do not observe. For instance, if low-performing firms are more likely to implement significant changes that require time to show positive effects, our results would underestimate benefits for this group. However, if the “generic” advice implemented by low performers (e.g., price cuts and advertising investments) proves counterproductive over longer horizons, our analysis would underestimate negative impacts. Our experiment also took place from May to November 2023, when exposure to and experience with generative AI tools such as ChatGPT was still low, particularly in developing market contexts such as Kenya. As business owners and workers become more familiar with generative AI tools, experiments such as ours may yield different results because treated participants, control participants, and even competing businesses outside the experiment may all be using AI tools.

Furthermore, our study evaluates an AI assistance built on GPT-4 with only a relatively short system prompt and that was distributed to entrepreneurs with 5-10 minutes of online training. Although substantial testing and iteration went into the developing our system prompt, it is possible that fine-tuning the language model using a corpus of relevant information, further iterating on the system prompt, and/or providing more training to entrepreneurs interacting with the AI assistant could improve the suggested positive impacts among high performers and reduce (or reverse) the adverse effects among low performers. Such improvements should be feasible given that interactions with the AI assistant are fully observed, allowing qualitative and quantitative analysis of questions asked, topics discussed, and answers provided (Choi et al., 2023). Additionally, the AI assistant’s heterogeneous impacts may be partially attributable to biases in GPT-4’s training data, and a model trained on better data could yield positive returns to AI for even the most challenging business tasks (Tao et al., 2023).

Finally, while our abductive mechanism analysis is most consistent with variation in judgment and ability driving differences in the types of AI advice that is implemented, we cannot fully rule out alternative possibilities. As with many studies in emerging markets (De Mel, McKenzie, and Woodruff, 2009), we also rely on self-reported survey data. As a consequence, our findings could partially reflect differences in how high and low perform-

ers respond to treatment in their reporting behavior, rather than actual business changes. Although we attempted to mitigate some of these concerns by using a control group that received business training materials (rather than no intervention), we cannot completely rule out reporting-based explanations for our findings given the exploratory nature of our mechanism analyses. Similarly, while our study’s sample size of 640 entrepreneurs is comparable to other experiments with firms in emerging markets (McKenzie, 2021), like these other studies, we still face challenges when it comes to statistical power (McKenzie, 2025). As a consequence, while high performers appear to make more tailored business changes and might have benefited from using the AI, we cannot strongly rule out the possibility that these patterns could result from sampling variability.

6.2 Implications

Our findings, along with the limitations discussed above, suggest several opportunities for further research and policy consideration. First, we need a better understanding of the mechanisms that drive entrepreneurs and workers to select and implement different types of advice. Although we find that low-performing entrepreneurs listen to and implement advice that harms performance, we do not know whether we can improve their choices, whether the negative effects stem from more fundamental differences in skills or capabilities, or why some participants continue to use AI despite experiencing negative returns during our study. The strong correlation and causal relationship between AI advice content and subsequent business changes (Table 4) suggests that entrepreneurs are highly responsive to AI recommendations, which underscores the critical importance of ensuring high-quality, contextually relevant advice before deploying such technologies. This responsiveness implies that poor or generic advice can easily lead business owners astray, as our results demonstrate with low-performing entrepreneurs. Even for high-performing entrepreneurs, is there room to further improve how they use the AI assistant or is the impact of AI advice already at the frontier of returns we should expect from this new technology (Doshi and Hauser, 2023)? One possibility is that it may be effective to provide bundles of complementary resources alongside AI advice, as the ability to act on good recommendations may depend not only on judgment but also on access to flexible capital and other implementation resources. Future work might also

explore both whether simple prompt modifications—such as instructing AI systems to avoid generic recommendations like discounting or broad advertising—help mitigate the negative effects we observe among lower-performing entrepreneurs (though this approach may simply shift users toward other suboptimal advice), and how learning dynamics and behavioral adaptation evolve over time, as our results suggest considerable variation in entrepreneurs' AI usage patterns during extended exposure periods.

Turning to our setting, our findings with Kenyan SMB entrepreneurs have important implications for AI deployment in developing economies. The heterogeneous effects we document suggest that untargeted AI deployment could exacerbate existing performance gaps between high and low performers, potentially widening gaps in business outcomes. This has particular relevance for policy makers and organizations seeking to use AI to promote economic development, as our results suggest that without careful targeting or complementary training, AI tools may benefit those who are already succeeding while further disadvantaging those who are struggling. Yet, the fact that the majority of entrepreneurs appear to follow at least some of the AI's advice suggests that AI assistance has the potential to benefit millions if not billions of workers and firms in developing economies. Akin to how the business training literature has identified that some types of training are effective whereas others seem to have little to no impact ([McKenzie, 2021](#)), researchers should further explore how we can improve the ways we deploy, design, and scale AI systems to drive inclusive growth in emerging markets ([George, McGahan, and Prabhu, 2012](#)). One promising avenue may be developing more focused AI systems that target specific domains—such as agricultural yields, market research, or financial practices—rather than providing broad, general-purpose advice that requires extensive filtering and prioritization.

It is also crucial to understand how our results generalize beyond developing economies. The underlying mechanism most consistent with our findings—that AI's impact depends critically on users' ability to exercise judgment in selecting and implementing advice—likely applies across many contexts where AI provides open-ended support rather than performing narrowly defined tasks. As AI's capabilities improve, AI will both provide better assistance on average and will be applied to a broader set of tasks, opening up the possibility that as AI improves, judgment may counterintuitively become even more important. In either case,

given our findings that human judgment and ability appear to strongly moderate the impact of AI, it will be increasingly important to understand how entrepreneurs, workers, and firms shift the skills they invest in and the ideas they pursue (Lebovitz, Lifshitz-Assaf, and Levina, 2022; Boussioux et al., 2023; Girotra et al., 2023; Hui, Reshef, and Zhou, 2023). Beyond judgment, effective AI use will continue to depend on users' ability to communicate relevant contextual details and adapt to evolving model capabilities, as recent research suggests that optimal prompting strategies may be model-specific (Jahani et al., 2024). In general, as AI changes, people and firms will too, potentially leading to unexpected equilibrium outcomes as both AI and humans learn.

Finally, our results demonstrate the value of studying AI's impact on the open-ended, day-to-day decisions that define real business operations (Harrison and List, 2004). By examining entrepreneurship in a developing economy context, our experiment reveals how AI's effects differ markedly from those observed in controlled, task-specific settings. This work underscores the critical need to evaluate AI across diverse contexts and populations—not merely to understand its economic potential, but to ensure that as this transformative technology reshapes work and business globally, it creates opportunities rather than deepening existing performance gaps.

References

- Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. 2023. “Combining human expertise with artificial intelligence: Experimental evidence from radiology.” Tech. rep., National Bureau of Economic Research.
- Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2022. *Prediction machines, updated and expanded: The simple economics of artificial intelligence*. Harvard Business Press.
- Agrawal, Ajay, Joshua S Gans, and Avi Goldfarb. 2023. “Do we want less automation?” *Science* 381 (6654):155–158.
- Agrawal, Ajay, Joshua S Gans, and Scott Stern. 2021. “Enabling entrepreneurial choice.” *Management Science* 67 (9):5510–5524.
- Anderson, Stephen J and David McKenzie. 2022. “Improving business practices and the boundary of the entrepreneur: A randomized experiment comparing training, consulting, insourcing, and outsourcing.” *Journal of Political Economy* 130 (1):157–209.
- Asiedu, Edward, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry. 2021. “A call for structured ethics appendices in social science papers.” *Proceedings of the National Academy of Sciences* 118 (29):e2024570118.
- Autor, David. 2014. “Polanyi’s paradox and the shape of employment growth.” NBER Working Paper 20485.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. “High-dimensional methods and inference on structural and treatment effects.” *Journal of Economic Perspectives* 28 (2):29–50.
- Benjamini, Yoav and Yosef Hochberg. 1995. “Controlling the false discovery rate: a practical and powerful approach to multiple testing.” *Journal of the Royal statistical society: series B (Methodological)* 57 (1):289–300.
- Björkegren, Daniel. 2023. “Artificial Intelligence for the Poor: How to Harness the Power of AI in the Developing World.” Foreign Affairs.
- Boussioux, Leonard, Jacqueline N Lane, Miaomiao Zhang, Vladimir Jacimovic, and Karim R Lakhani. 2023. “The Crowdless Future? How Generative AI Is Shaping the Future of Human Crowdsourcing.” *The Crowdless Future* .
- Breiman, Leo. 2001. “Random forests.” *Machine learning* 45 (1):5–32.
- Bryan, Kevin A, András Tilcsik, and Brooklynn Zhu. 2017. “Which entrepreneurs are coachable and why?” *American Economic Review* 107 (5):312–316.
- Brynjolfsson, Erik and Lorin M Hitt. 2000. “Beyond computation: Information technology, organizational transformation and business performance.” *Journal of Economic Perspectives* 14 (4):23–48.
- Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. “Generative AI at work.” *The Quarterly Journal of Economics* :qjae044.

- Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond. 2023. “Generative AI at work.” NBER Working Paper 31161.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. “The productivity J-curve: How intangibles complement general purpose technologies.” *American Economic Journal: Macroeconomics* 13 (1):333–372.
- Callen, Michael, Miguel Fajardo-Steinhäuser, Michael G Findley, and Tarek Ghani. 2025. “Can digital aid deliver during humanitarian crises?” *Management Science* .
- Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C Johnson, David McKenzie, and Mona Mensmann. 2017. “Teaching personal initiative beats traditional training in boosting small business in West Africa.” *Science* 357 (6357):1287–1290.
- 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.
- Cao, Sam Ruiqing, Rembrand Koning, and Ramana Nanda. 2023. “Sampling Bias in Entrepreneurial Experiments.” *Management Science*.
- Carlson, Natalie A. 2023. “Differentiation in microenterprises.” *Strategic Management Journal* 44 (5):1141–1167.
- Chandler, Alfred D. 1977. *The visible hand*. Harvard university press.
- Chatterji, Aaron, Solène Delecourt, Sharique Hasan, and Rembrand Koning. 2019. “When does advice impact startup performance?” *Strategic Management Journal* 40 (3):331–356.
- Chen, Jiafeng and Jonathan Roth. 2024. “Logs with zeros? Some problems and solutions.” *The Quarterly Journal of Economics* 139 (2):891–936.
- Chen, Zenan and Jason Chan. 2023. “Large language model in creative work: The role of collaboration modality and user expertise.” Available at SSRN 4575598 .
- Choi, Jun Ho, Oliver Garrod, Paul Atherton, Andrew Joyce-Gibbons, Miriam Mason-Sesay, and Daniel Björkegren. 2023. “Are LLMs Useful in the Poorest Schools? theTeacherAI in Sierra Leone.” *arXiv preprint arXiv:2310.02982* .
- Cusolito, Ana P, Ornella Darova, and David McKenzie. 2023. “Capacity building as a route to export market expansion: A six-country experiment in the Western Balkans.” *Journal of International Economics* 144:103794.
- Davies, Elwyn, Peter Defebach, Leonardo Iacovone, and David Mckenzie. 2024. “Training microentrepreneurs over Zoom: Experimental evidence from Mexico.” *Journal of Development Economics* 167:103244.
- De Mel, Suresh, David J McKenzie, and Christopher Woodruff. 2009. “Measuring microenterprise profits: Must we ask how the sausage is made?” *Journal of development Economics* 88 (1):19–31.

- Dell'Acqua, Fabrizio, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Cadelon, 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.
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. “Bert: Pre-training of deep bidirectional transformers for language understanding.” *arXiv preprint arXiv:1810.04805* .
- 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.
- . 2024. “Networking Frictions: Evidence from entrepreneurial networking events in Lomé.” *Management Science* (forthcoming).
- Doshi, Anil R and Oliver Hauser. 2023. “Generative artificial intelligence enhances creativity.” Available at SSRN 4535536.
- Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. “GPTs are GPTs: Labor market impact potential of LLMs.” *Science*.
- Fabregas, Raissa, Michael Kremer, Matthew Lowes, Robert On, and Giulia Zane. 2024. “Digital information provision and behavior change: Lessons from six experiments in East Africa.” Tech. rep., National Bureau of Economic Research.
- George, Gerard, Anita M McGahan, and Jaideep Prabhu. 2012. “Innovation for inclusive growth: Towards a theoretical framework and a research agenda.” *Journal of management studies* 49 (4):661–683.
- Girotra, Karan, Lennart Meincke, Christian Terwiesch, and Karl T Ulrich. 2023. “Ideas are dimes a dozen: Large language models for idea generation in innovation.” Available at SSRN 4526071.
- Glennerster, Rachel and Kudzai Takavarasha. 2014. *Running randomized evaluations: A practical guide*. Princeton University Press.
- Goldberg, Amir and Sameer B Srivastava. 2022. “How Can AI Enrich Our Understanding of Organizational Culture.”
- Goldfarb, Avi. 2024. “Pause artificial intelligence research? Understanding AI policy challenges.” *Canadian Journal of Economics/Revue canadienne d'économique* .
- Grimmer, Justin and Brandon M Stewart. 2013. “Text as data: The promise and pitfalls of automatic content analysis methods for political texts.” *Political analysis* 21 (3):267–297.
- Guzman, Jorge and Aishen Li. 2023. “Measuring founding strategy.” *Management Science* 69 (1):101–118.
- Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. “Learning through noticing: Theory and evidence from a field experiment.” *The Quarterly Journal of Economics* 129 (3):1311–1353.

- Harrison, Glenn W and John A List. 2004. “Field experiments.” *Journal of Economic Literature* 42 (4):1009–1055.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2nd ed.
- Holtz, David, Felipe Lobel, Ruben Lobel, Inessa Liskovich, and Sinan Aral. 2023. “Reducing interference bias in online marketplace experiments using cluster randomization: Evidence from a pricing meta-experiment on airbnb.” *Management Science* .
- Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2023. “The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market.” SSRN 4527336.
- ILO. 2014. *Start and Improve Your Business (SIYB)*. Geneva. URL https://www.ilo.org/empent/areas/start-and-improve-your-business/WCMS_192062/lang--en/index.htm. Accessed: 2025-05-18.
- Jahani, Eaman, Benjamin S Manning, Joe Zhang, Hong-Yi TuYe, Mohammed Alsobay, Christos Nicolaides, Siddharth Suri, and David Holtz. 2024. “As generative models improve, we must adapt our prompts.” *University of California, Berkeley* .
- Ji, Ziwei, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. “Survey of hallucination in natural language generation.” *ACM Computing Surveys* 55 (12):1–38.
- Kamau, Paul, Laura Barasa, Gedion Onyango, and Oscar Otele. 2024. “Summary of results: Afrobarometer Round 10 survey in Kenya, 2024.” Survey summary, Institute for Development Studies, University of Nairobi & Afrobarometer. URL <https://www.afrobarometer.org/wp-content/uploads/2024/10/Summary-of-results-Kenya-Afrobarometer-R10-template-Eng-16oct24-for-upload.pdf>.
- Karlan, Dean, Ryan Knight, and Christopher Udry. 2015. “Consulting and capital experiments with microenterprise tailors in Ghana.” *Journal of Economic Behavior & Organization* 118:281–302.
- Keen, Peter GW and Michael S Scott Morton. 1978. “Decision support systems: an organizational perspective.” (*No Title*) .
- Kenya National Bureau of Statistics. 2023. “Economic Survey 2023: Popular Version.” URL <https://www.knbs.or.ke/wp-content/uploads/2023/09/2023-Economic-Survey-Popular-Version.pdf>.
- Khanna, Tarun. 2014. “Entrepreneurship in emerging markets: Contextual intelligence for the study of two thirds of the world’s population.” In *Multidisciplinary insights from new AIB fellows*, vol. 16. Emerald Group Publishing Limited, 221–238.
- 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.

- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. “Experimental analysis of neighborhood effects.” *Econometrica* 75 (1):83–119.
- Kuhn, Max. 2008. “Building predictive models in R using the caret package.” *Journal of statistical software* 28 (5):1–26.
- Kumar, Harsh, David M Rothschild, Daniel G Goldstein, and Jake M Hofman. 2023. “Math Education with Large Language Models: Peril or Promise?” Available at SSRN 4641653.
- Lazear, Edward P. 2004. “Balanced skills and entrepreneurship.” *American Economic Review* 94 (2):208–211.
- Lebovitz, Sarah, Hila Lifshitz-Assaf, and Natalia Levina. 2022. “To engage or not to engage with AI for critical judgments: How professionals deal with opacity when using AI for medical diagnosis.” *Organization science* 33 (1):126–148.
- Lou, Bowen, Hongshen Sun, and Tianshu Sun. 2023. “GPTs and labor markets in the developing economy: Evidence from China.” Available at SSRN 4426461 .
- McAfee, A, D Rock, and E Brynjolfsson. 2023. “How to Capitalize on Generative AI.” *Harvard Business Review* 101 (6):42–48.
- McKenzie, David. 2012. “Beyond baseline and follow-up: The case for more T in experiments.” *Journal of Development Economics* 99 (2):210–221.
- . 2021. “Small business training to improve management practices in developing countries: re-assessing the evidence for ‘training doesn’t work’.” *Oxford Review of Economic Policy* 37 (2):276–301.
- . 2023. “Interpreting treatment effects on an inverse hyperbolic sine outcome variable and alternatives.” *Development Impact* Accessed: 2023-11-29.
- . 2025. “Designing and analysing powerful experiments: practical tips for applied researchers.” *Fiscal Studies* .
- McKenzie, David, Adam Osman, and Aminur Rahman. 2023. “Training and subsidies vs pay-for-results in spurring digital marketing take-up and small firm growth.” .
- McKenzie, David and Christopher Woodruff. 2017. “Business practices in small firms in developing countries.” *Management Science* 63 (9):2967–2981.
- Mehmood, Muhammad Zia. 2023. “Short Messages Fall Short for Micro-Entrepreneurs: Experimental Evidence from Kenya.” .
- Meincke, Lennart, Karan Girotra, Gideon Nave, Christian Terwiesch, and Karl T Ulrich. 2024. “Using large language models for idea generation in innovation.” *The Wharton School Research Paper Forthcoming* 9:2024.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. “Efficient estimation of word representations in vector space.” *arXiv preprint arXiv:1301.3781* .
- Mollick, Ethan. 2022. “ChatGPT is a tipping point for AI.” *Harvard Business Review* .

- Mollick, Ethan R and Lilach Mollick. 2022. “New modes of learning enabled by ai chatbots: Three methods and assignments.” *Available at SSRN 4300783* .
- Noy, Shaked and Whitney Zhang. 2023. “Experimental evidence on the productivity effects of generative artificial intelligence.” *Science* 381:187–192.
- OpenAI. 2023. “GPT-4 Technical Report.”
- OpenAI. 2024. “New embedding models and API updates.” URL <https://openai.com/index/new-embedding-models-and-api-updates/>. Accessed: October 3, 2025.
- Pearl, Judea. 2009. *Causality*. Cambridge university press.
- Peng, Sida, Eirini Kalliamvakou, Peter Cihon, and Mert Demirer. 2023. “The Impact of AI on Developer Productivity: Evidence from Github Copilot.” *arXiv preprint:2302.06590* .
- Pennington, Jeffrey, Richard Socher, and Christopher D Manning. 2014. “Glove: Global vectors for word representation.” In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.
- Pillai, Sandeep Devanatha, Brent Goldfarb, and David Kirsch. 2024. “Lovely and likely: Using historical methods to improve inference to the best explanation in strategy.” *Strategic Management Journal* 45 (8):1539–1566.
- Radford, Alec, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever et al. 2019. “Language models are unsupervised multitask learners.” *OpenAI blog* 1 (8):9.
- Safaricom PLC. 2024. “Safaricom’s M-PESA Hits 34 Million Customers in Kenya.” URL <https://www.safaricom.co.ke/media-center-landing/press-releases/safaricoms-m-pesa-hits-34-million-customers-in-kenya>.
- Sato, Megan Kinniment Lucas Jun Koba, Haoxing Du, Brian Goodrich, Max Hasin, Lawrence Chan, Luke Harold Miles, Tao R Lin, Hjalmar Wijk, Joel Burget, Aaron Ho, Elizabeth Barnes, and Paul Christiano. 2023. “Evaluating Language-Model Agents on Realistic Autonomous Tasks.” arXiv: 2312.11671.
- Sharda, Ramesh, Steve H Barr, and James C McDonnell. 1988. “Decision support system effectiveness: a review and an empirical test.” *Management science* 34 (2):139–159.
- Shengale, Ram. 2025. “WhatsApp Statistics 2025: Usage Trends, Demographics & More.” URL <https://wanotifier.com/whatsapp-statistics/>.
- Sprague Jr, Ralph H. 1980. “A framework for the development of decision support systems.” *MIS quarterly* :1–26.
- Tao, Yan, Olga Viberg, Ryan S. Baker, and Rene F. Kizilcec. 2023. “Auditing and Mitigating Cultural Bias in LLMs.” arXiv: 2311.14096.
- The White House. 2022. ““The Impact of Artificial Intelligence on the Future of Workforces in the European Union and the United States of America.” Technical Report.

The World Bank Group. 2018. “Enterprise Surveys: What Businesses Experience - Kenya 2018 Country Profile.” URL <https://www.enterprisesurveys.org/content/dam/enterprisesurveys/documents/country-profiles/Kenya-2018.pdf>.

Wamuyu, PK. 2020. “The Kenyan Social Media Landscape: Trends and Emerging Narratives, 2020.” SIMElab Working Paper.

Wiles, Emma and John J Horton. 2024. “More, but Worse: The Impact of AI Writing Assistance on the Supply and Quality of Job Posts.” .

Zizzo, Daniel John. 2010. “Experimenter demand effects in economic experiments.” *Experimental Economics* 13:75–98.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean S Karlan, Richard Hornbeck, Xavier Giné et al. 2011. “Being surveyed can change later behavior and related parameter estimates.” *Proceedings of the National Academy of Sciences* 108 (5):1821–1826.

Figures

Figure 1: Overview of the experimental timeline from the beginning of sample recruitment to endline survey.

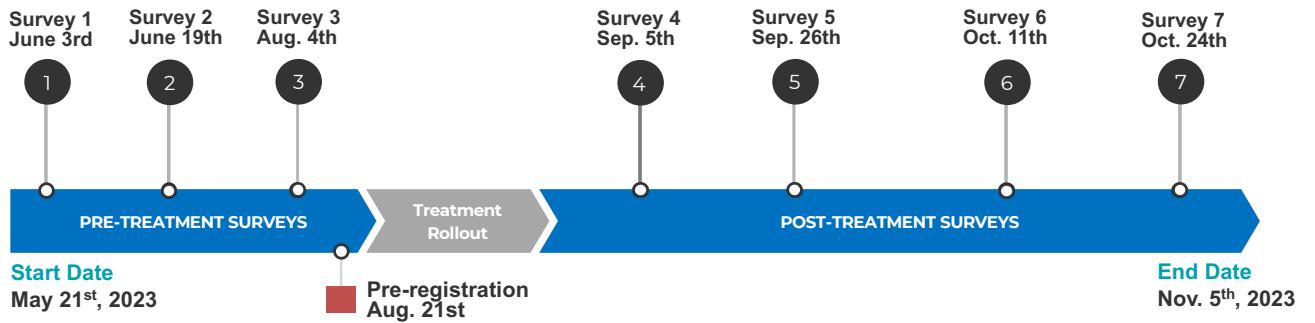


Figure 2: Example of an Entrepreneur-AI interaction over WhatsApp.

I have fast food joint that I run just adjacent to a matatu terminus. There are several such food joints at that particular place and it's fair to say competition is stiff. How can I overcome the competitive business environment and be ahead of the rest of the pack?

Thank you for sharing about your business. Operating a fast food joint in a highly competitive environment like a matatu terminus can indeed be challenging, but there are several strategies you could consider to stand out:

1. **Differentiate Your Menu:** Offer unique or special dishes that aren't available at the other joints. This could be traditional Kenyan foods with a twist, or it could involve using locally sourced ingredients that others don't use. The benefit is that customers will associate your place with these special offerings and choose you over competitors.
2. **Improve Customer Service:** Prioritize excellent customer service to make people feel welcome and appreciated when they visit your joint. This might involve training staff on how to interact with customers positively, quickly serving orders, and handling any complaints promptly and professionally.
3. **Loyalty Program:** Implement a loyalty program where customers earn points for every purchase which they can later redeem for free meals or discounts. It encourages repeat business as customers will want to accumulate more points.
4. **Community Involvement:** Engage in local community events or sponsor local initiatives such as sports teams or school activities. This raises your profile within the community and makes people more likely to support your business.
5. **Cleanliness & Hygiene:** Ensure high standards of cleanliness both inside the kitchen and in the dining area because hygiene matters greatly in food businesses.

Enter a number to learn more about one of these points on "improving competitiveness" or ask a new question.

3

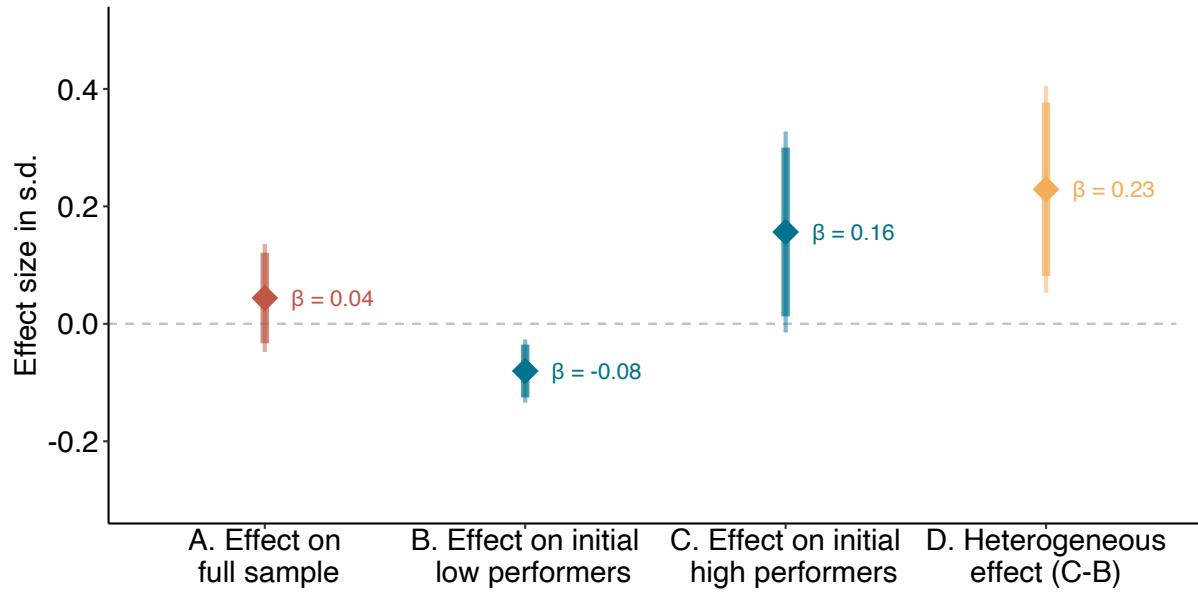
Great, let's delve into how you can implement a **Loyalty Program** in your fast food joint:

1. **Point-Based System:** Implement a simple point-based system where customers earn points for every shilling spent. Once they accumulate a certain number of points, they can redeem them for free meals or discounts. You could use simple tools like punch cards or digital apps that are inexpensive to manage.
2. **Special Offers for Regular Customers:** Identify your regular customers and offer them exclusive deals or discounts. This not only rewards their loyalty but also encourages them to keep coming back.
3. **Birthday/Anniversary Rewards:** Offer special treats or discounts to customers on their birthdays or anniversaries if they share this information with you. This personalized touch can make customers feel valued and appreciated.
4. **Referral Program:** Encourage your loyal customers to bring in new clients by giving them incentives such as a discount or free meal when someone they refer makes their first purchase.
5. **Feedback Reward:** Encourage customer feedback by offering small rewards for completing surveys about the quality of food and service at your joint.

Remember, the goal of any loyalty program is to make your customers feel valued so that they continue patronizing your business over others. Enter a number to learn more about one of these points on "loyalty programs" or ask a new question.

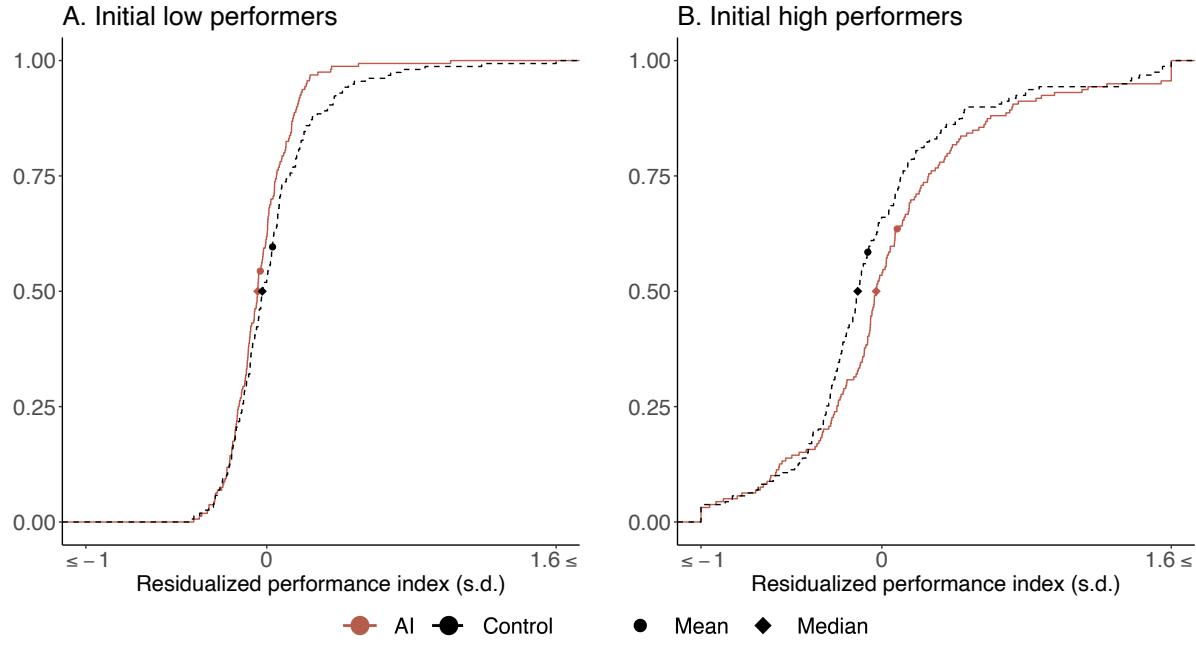
This figure presents a rendering of an entrepreneur-AI interaction in WhatsApp. The text is original (including the typos). A "matatu" refers to a minibus in Kenya.

Figure 3: Effects of the AI Assistant on Business Performance



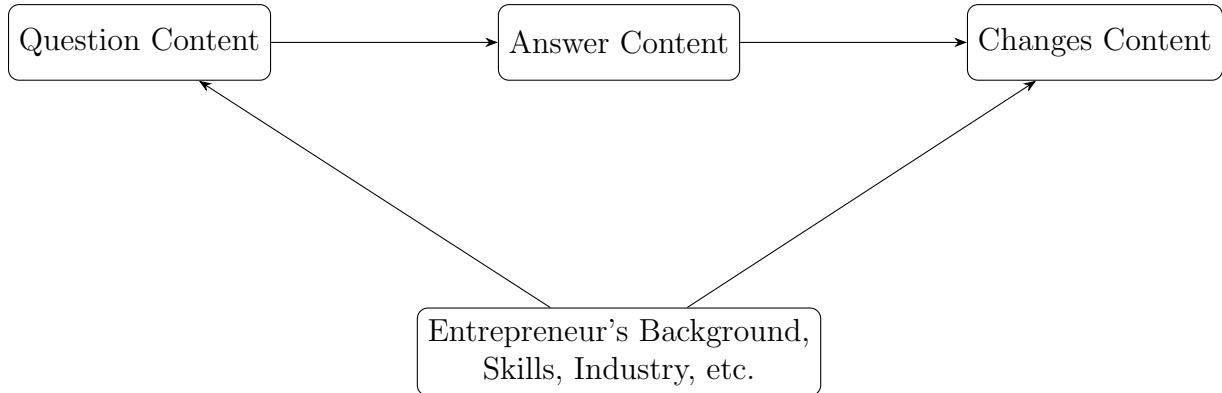
The impact of access to the AI assistant on the standardized business performance index. Effects are estimated controlling for pre-treatment business performance, time fixed effects, stratum fixed effects, and other covariates selected using double-LASSO. Estimate A presents the average treatment effect of the AI assistant. Estimates B and C display results restricted to initially low- and high-performing firms. Estimate D presents the heterogeneous treatment effect. Error bars represent 90% and 95% confidence intervals, with standard errors clustered at the individual level.

Figure 4: Empirical Cumulative Distribution Functions of Residualized Business Performance



Panels A and B display empirical cumulative distribution functions (ECDFs) of the standardized business performance index. Red solid lines represent the AI assistant condition, and black dashed lines denote the control group. Panel A presents results for participants with below-median pre-treatment performance, while Panel B shows those with above-median pre-treatment performance. Higher values on the x -axis correspond to higher levels of business performance.

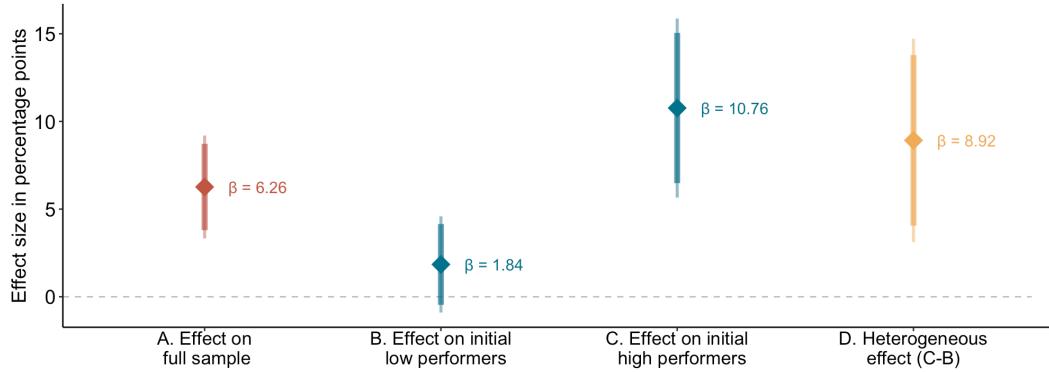
Figure 5: Directed Acyclic Graph Illustrating the Backdoor Criterion



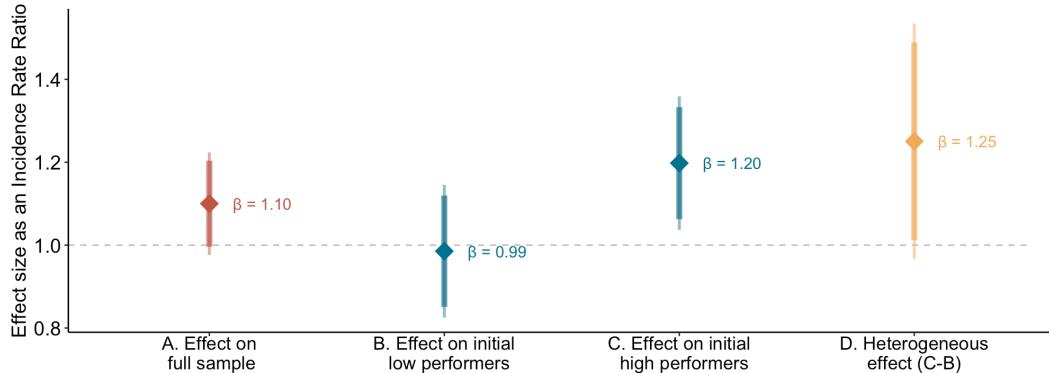
Entrepreneur characteristics influence both the questions asked and business changes made, creating potential confounding. Since the AI assistant only accesses information through WhatsApp messages, controlling for question content blocks all backdoor paths between AI advice and business changes, enabling causal identification of the AI's impact on entrepreneurial decisions.

Figure 6: The Heterogeneous Impact of the AI Treatment on the Entrepreneurs' Business Changes

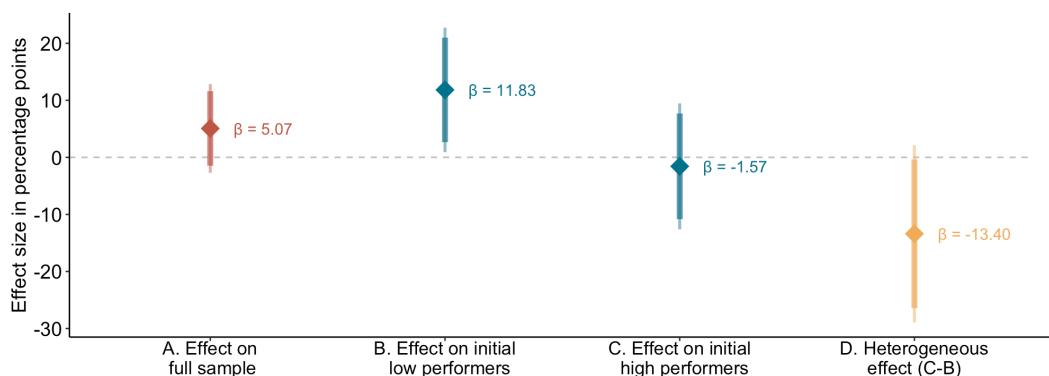
Panel A: Describes Working and Learning from the AI?



Panel B: Count of Uncommon Words

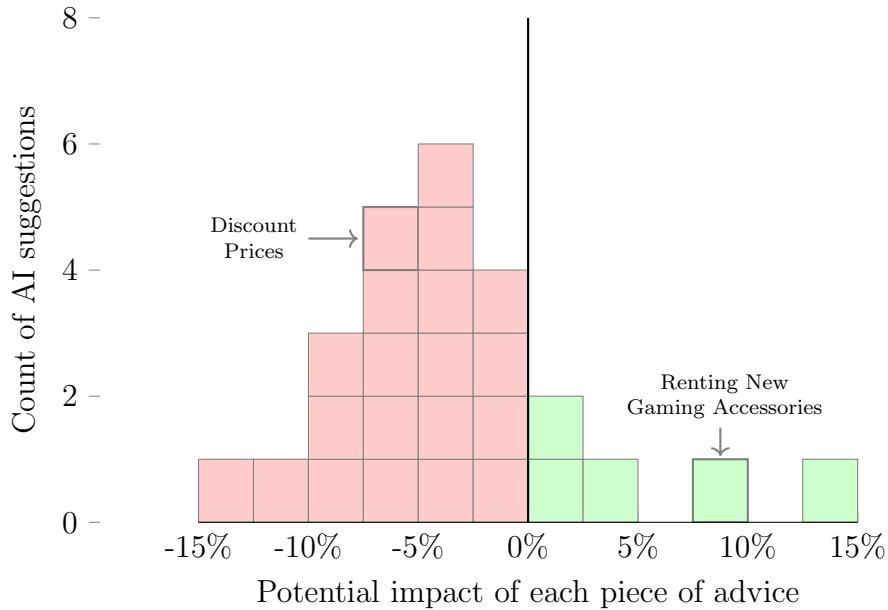


Panel C: Mentions Discounting and Advertising?



These panels quantitatively test the qualitative results in [Table 5](#). Panel A shows that high performers are more likely to describe working and learning from the AI assistant. Panel B shows that high performers use more uncommon words that are tailored to their specific business needs. Panel C shows that low performers are more likely to implement generic and potentially costly strategies like lowering prices or investing in ads. Error bars present 90 and 95% confidence intervals. Regression at the entrepreneur level with robust standard errors. The estimates in Panel A are from the linear probability models in [Table A7](#), the estimates in Panel B are the exponentiated coefficients from the Poisson regressions in [Table A8](#), and the estimates in Panel C from the linear probability models in [Table A7](#).

Figure 7: Illustration of Proposed Selection and Implementation Mechanism



This stylized figure illustrates how heterogeneous treatment effects can arise when entrepreneurs have different abilities to select and implement advice from a mixed-quality pool of AI suggestions. When some AI advice reduces performance (such as generic price-cutting strategies) while other suggestions meaningfully improve it (such as targeted product additions), variation in entrepreneurs' selection and implementation choices produces large outcome differences. Prior generative AI experiments have largely focused on well-defined tasks where suggestions are uniformly beneficial or harmful (e.g., [Brynjolfsson, Li, and Raymond, 2025](#); [Dell'Acqua et al., 2023](#)), reducing the distribution of advice quality and minimizing the role of user judgment in determining outcomes.

Tables

Table 1: Treatment Effects of AI Assistant Access: Average Effects and Heterogeneity by Pre-Treatment Performance

	(A) Full Sample	(B) Below Median	(C) Above Median	(D) Full Sample
AI	0.04 (0.05)	-0.08** (0.03)	0.16† (0.09)	-0.07* (0.03)
AI \times (\geq med.)				0.23* (0.09)
Observations	2,514	1,251	1,263	2,514

The impact of access to the AI assistant on firm performance. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Columns A and D report effects across the full sample. Columns B and C report effects restricting to participants with above- or below-median pre-treatment performance. Effects are estimated controlling for pre-treatment performance, time fixed effects, stratum fixed effects, and covariates selected using double-LASSO. The outcome is winsorized at the 99% level. Standard errors are clustered at the individual level.

Table 2: Text Data Availability and Word Counts by Treatment Status and Performance Level

	Changes Text		Questions Text		Answers Text	
	Has Text?	Num. Word	Has Text?	Num. Words	Has Text?	Num. Words
	(1) OLS	(2) Poisson	(3) OLS	(4) Poisson	(5) OLS	(6) Poisson
Constant	0.8205*** (0.0308)	3.831*** (0.0778)	0.8820*** (0.0255)	4.662*** (0.1336)	0.8882*** (0.0249)	7.916*** (0.1827)
AI	0.0304 (0.0418)	-0.0274 (0.1083)				
(\geq Median)	0.0411 (0.0413)	0.1510 (0.1115)	-0.0402 (0.0387)	0.0813 (0.1857)	-0.0464 (0.0383)	-0.2770 (0.2163)
AI \times (\geq Median)	-0.0376 (0.0574)	0.0896 (0.1537)				
Subsample	All	All	AI	AI	AI	AI
Observations	634	634	319	319	319	319

Heteroskedasticity-robust standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Treatment status and baseline performance do not predict whether entrepreneurs have available text data or the length of their responses for business changes, questions asked, or AI responses received. This indicates that neither text availability nor response length is systematically related to treatment assignment or performance level, supporting the validity of text-based analyses.

Table 3: Predictive Accuracy of Text Embeddings for Entrepreneur Characteristics

Model	(1)	(2)	(3)	(4)	(5)
Data Prediction	<i>Questions</i> High Performer	<i>AI Advice</i> High Performer	<i>Changes</i> High Performer	<i>Changes</i> AI Treated	<i>Changes</i> AI Treated × High Performer
Classes Entrepreneurs	2 280	2 276	2 537	2 537	4 537
No Information Rate Accuracy	0.514 0.432	0.518 0.518	0.505 0.584	0.505 0.598	0.254 0.338
95% CI	[0.34, 0.53]	[0.42, 0.61]	[0.52, 0.65]	[0.53, 0.66]	[0.27, 0.41]
P-Value [Acc > NIR]	0.965	0.538	0.012	0.004	0.004

Random forests fit over high-dimensional word embeddings created from the questions asked and AI generated advice cannot predict whether an entrepreneur is a low or high performer. High-dimensional Word embeddings created from the self-reported business changes can predict whether an entrepreneur is a low or high performer and whether there receive the AI treatment or not. Four entrepreneurs asked a question of the AI, but due to technical errors, received no response.

Table 4: Effects of AI Advice Content on Business Changes

Model:	Changes Embedding Value					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.0005 (0.0107)	0.0005 (0.0086)				
Answer Embedding Value	0.2631*** (0.0156)	0.1274*** (0.0102)	0.1274*** (0.0102)	0.1266*** (0.0113)	0.1266*** (0.0113)	0.1280*** (0.0115)
Question Embedding Value		0.3450*** (0.0127)	0.3451*** (0.0127)	0.3320*** (0.0156)	0.3320*** (0.0156)	0.3588*** (0.0152)
Answer \times Above Median				0.0014 (0.0101)		
Question \times Above Median				0.0267 (0.0175)		
Performers Subsample	All	All	All	All	Low	High
<i>Fixed-effects</i>						
Entrepreneur			Yes	Yes	Yes	Yes
<i>Fit statistics</i>						
Observations	731,136	728,064	728,064	728,064	371,712	356,352

Clustered (Entrepreneur & Embedding Dimension) standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Regressions at the “embedding element”-entrepreneur level examine the relationship between AI advice content and business changes using embedding dimensions as observations. Model 1 shows the raw correlation (0.26 standard deviations), while Model 2 applies the backdoor criterion by controlling for question content, yielding a causal estimate of 0.13 standard deviations. The difference suggests that roughly half the correlation reflects confounding from entrepreneurs’ pre-existing plans rather than AI influence. Models 3-6 test robustness and show similar causal effects for both high and low performers, indicating that while entrepreneurs implement AI advice, the propensity to do so does not differ by performance level. [Figure 5](#) illustrates how controlling for question content enables causal identification by blocking confounding pathways.

Table 5: Example Business Changes with High Cosine Similarity to Treatment Effect Embedding

Control Examples	AI Examples
A. High Performers	
[...] I've started baking honey filled muffins. Honey baked muffins. Home delivery of my products. [...]	Use of new technology of AI mentor. By doing more benchmarking on the competitors. Cooking oil, cereals. [...]
Adding more products into the business. [...] Advertising what am selling. [...] Delivering door to door services.	Improved on customer care skills after getting some tips and advice from Almentor and introduction of discounts to customers who purchased the products in bulk. [...]
[...] I have started selling hair bonnets. [...] Delivery of products to customers has really made a difference in my business. Making myself available at any time for my customers. [...]	[...] I have inserted a small salon beside my cyber [...] I got advice from AI mentor on other ways of getting power when there is electricity power black out. [...]
[...] Cooking quality food. [...] Delivering my products to customers. Advertising my business through online. [...] Giving my customers carrier bags. [...] Selling coffee	Improving the quality of chicks in my chicken roster has brought high quality chicken to my folder... I ma grateful for the Chat AI. [...] I have bought new chicks that were recommended by whe WhatsApp AI. [...] Reducing slightly the prices of my commodities [...]
[...] i introduced other kitchen ware sets like cups and plates. Delivery of goods to customers. Engaging clients. Use of media in advertising. Male clothes. [...]	[...] With the help of AI, I have been able to do more market research and discovered new ways of advertising my products. I have added cellphones, speakers and electric wires [...] I have started electronics repairs [...] I have started free delivery for Items worth 5000 and above.
B. Low Performers	
Increase in profit in the business. Sales of eggs. Marketing my goods online. Selling of clothes. Transport service. [...]	Giving so offers by reducing the selling price. [...] Advertising on social media platforms. [...]
[...] Getting to know what customers wants, precise record keeping and online marketing. [...] Reduced owing to reduce debts. [...] Providing loan at an interest.	Advertisement and discounts. Offering after sales service. Training of staff. [...]
Budgeting and data keeping. [...] Increased profits because I can now manage cashflow appropriately without possible losses. Maize. Mpesa services. Fruit selling. [...]	Posting online has brought more customers. [...] Discount for items above 1000. Offers for goods above 1500. [...] Selling shoes in my mtumba warehouse. [...] Cereal outside my shop
Consulting my customers. grocery. [...] using of technology. selling of cosmetics products. [...]	Isuing discounts on some products. Painting my shop as advertising strategy. Discount. Packaging.
[...] I have started selling materials for making vitenge. [...] Bodaboda service. [...] Including mitumbas in shoes selling. Selling credit. Started layering poultry	The free delivery for my products is really working out for me. [...] New Imperial leather for men. Hair dressing. [...] Rearing and selling chicken

Panel A displays results for initially high performing entrepreneurs while Panel B displays results for initially low performing entrepreneurs. [Table A9](#) and [Table A10](#) display extended text results. Highlights correspond to the qualitative differences we observe between the control and treated text.

Online Appendix

Table of Contents

A Appendix Figures	58
B Appendix Tables	69
C Details on the Experimental Interventions	80
C.1 Developing the Generative AI Assistant	80
C.2 Overview of Intervention Training	81
D Detailed Case Examples	83
E Survey Questions and Measurement	98
E.1 Measuring Business Performance	98
E.2 Explanation of Preferred Functional Forms	98
E.3 Survey-based Mechanisms Questions and Analysis	99
F Ethical Considerations	104
G Experimental Spillovers	105
H Text Analysis	106
H.1 Overview and Summary Statistics	107
H.2 Question and Answer Topic Analysis	108
H.3 Do Entrepreneurs Follow AI Advice?: Bag of Words Analysis	111
H.4 Technical Details: Treatment Effect Embedding Methodology	111
H.5 Quantitative Validation: Technical Implementation Details	112

A Appendix Figures

Figure A1: The Distribution of Study Participants Across Kenya

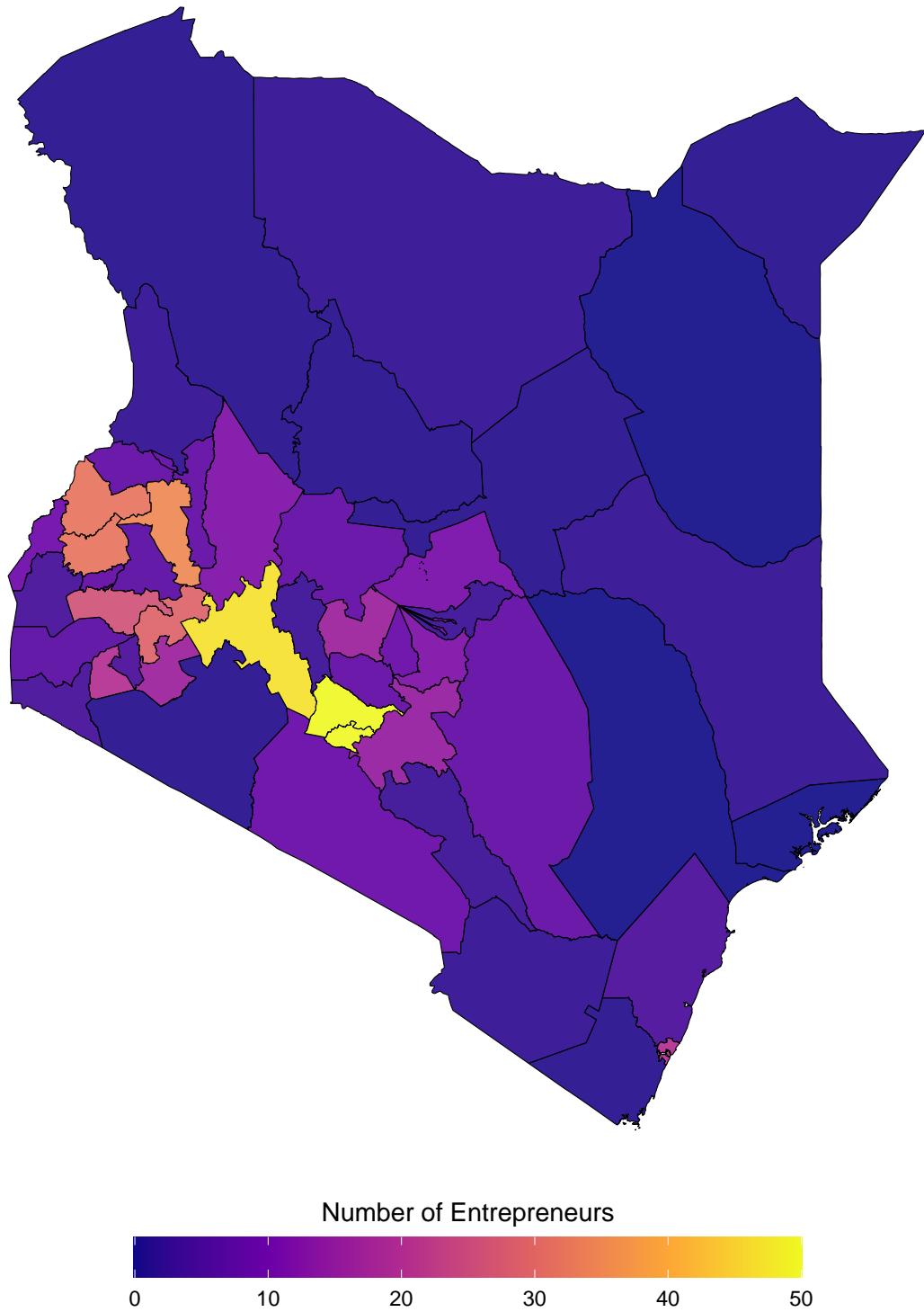
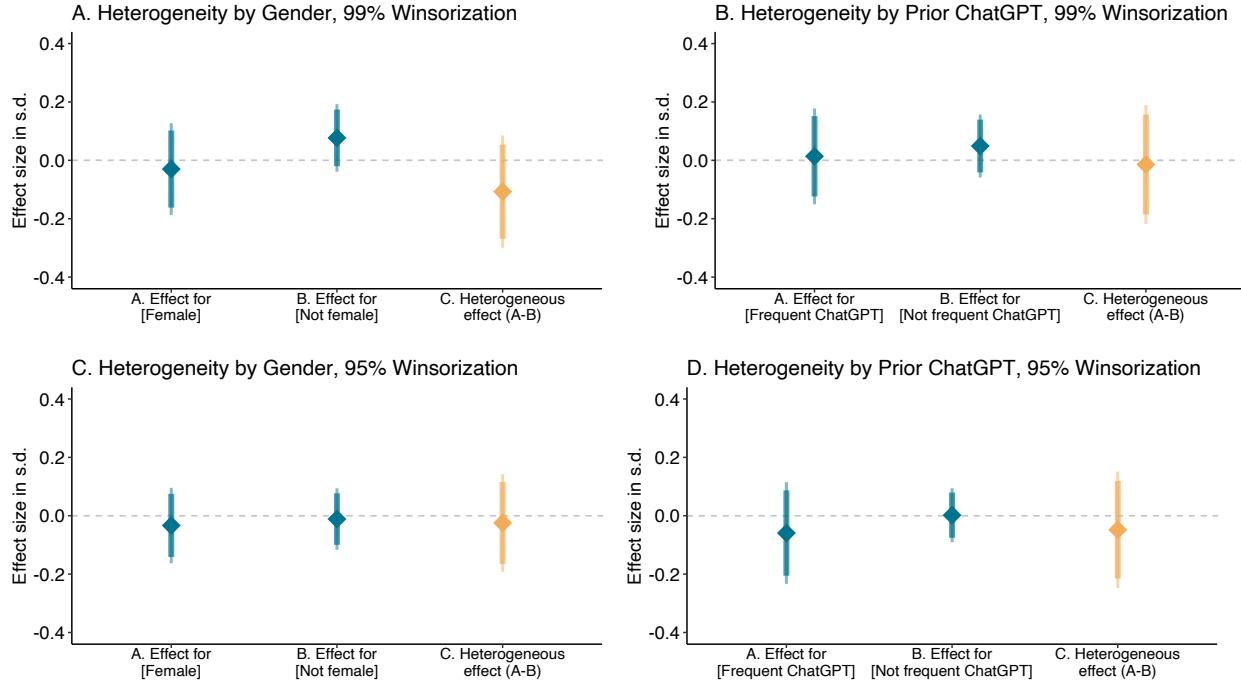
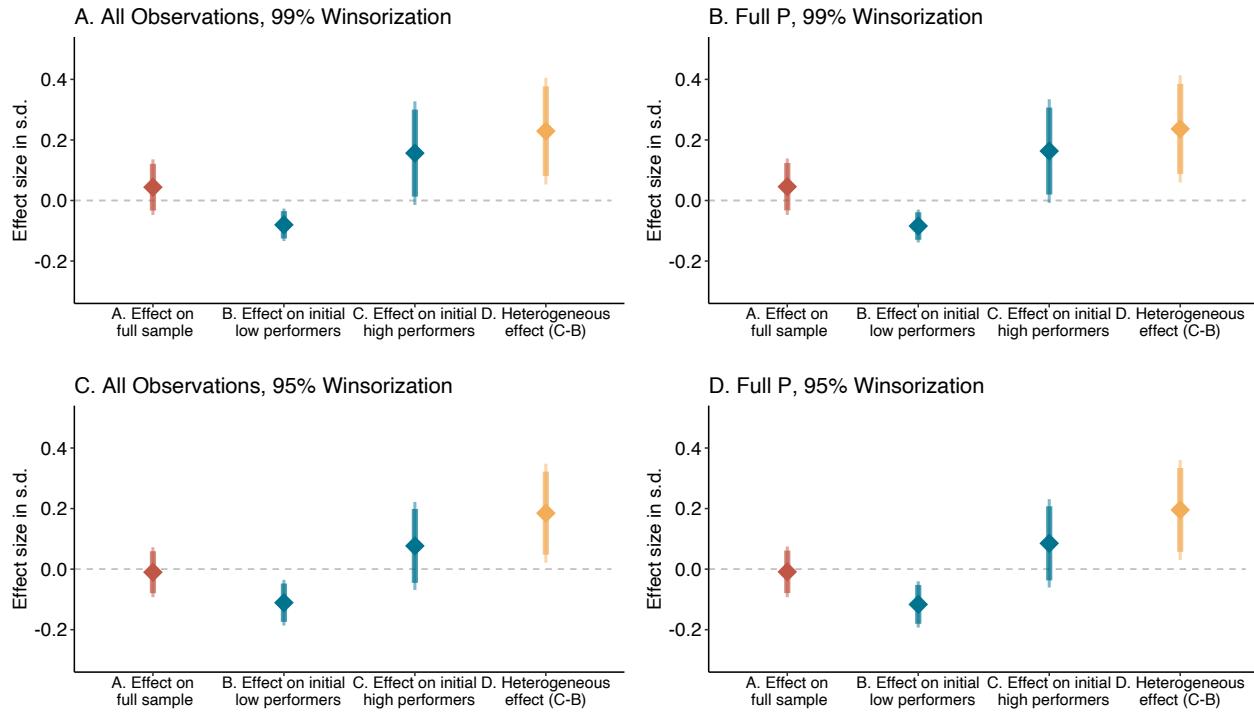


Figure A2: Performance effect heterogeneity by gender and prior ChatGPT use.



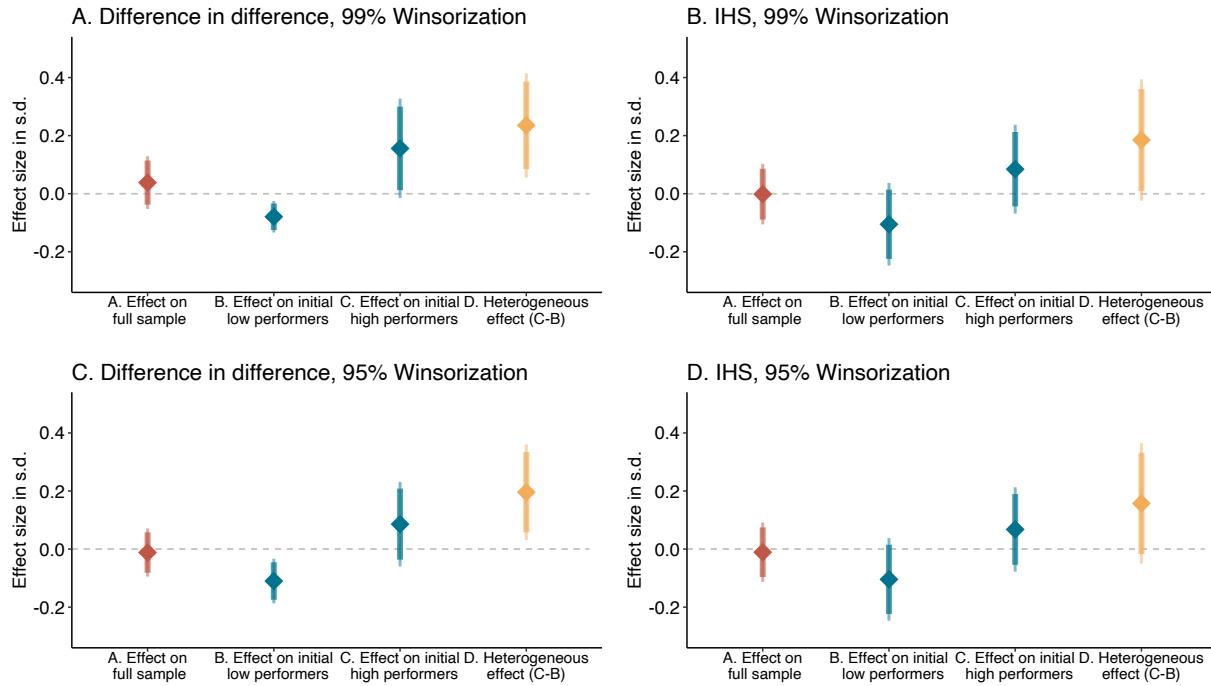
Heterogeneous effects of access to the AI assistant by gender and prior ChatGPT use. Panels A-B (C-D) display results winsorized at the 99% (95%) level. Panels A and C present heterogeneous effects based on participants' reported gender. Panels B and D report results based on prior ChatGPT use, where *frequent use* is defined as reporting using ChatGPT at least once per week. Effects are estimated controlling for covariates selected using double-LASSO, including pre-treatment business performance. Error bars represent 90% and 95% confidence intervals.

Figure A3: Tests of robustness to alternative sample inclusion criteria at both 95% and 99% winsorization.



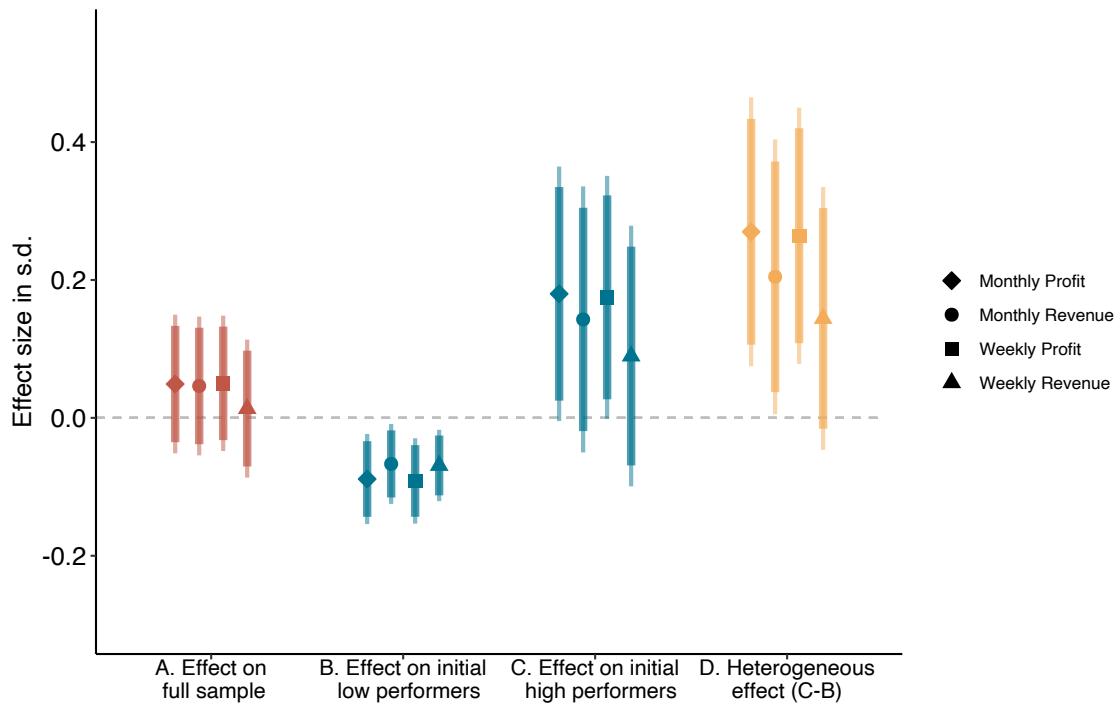
The effect of access to the AI assistant on the business performance index. Panels A and B (C and D) display results winsorized at the 99% (95%) level. Panels A and C include all observations, while Panels B and D restrict the analysis to participants completing all post-treatment surveys (i.e., “Full P”). Effects are estimated controlling for pre-treatment business performance, time fixed effects, stratum fixed effects, and covariates selected using double-LASSO. Error bars represent 90% and 95% confidence intervals.

Figure A4: Tests of robustness to alternative outcome variable transformations at both 95% and 99% winsorization.



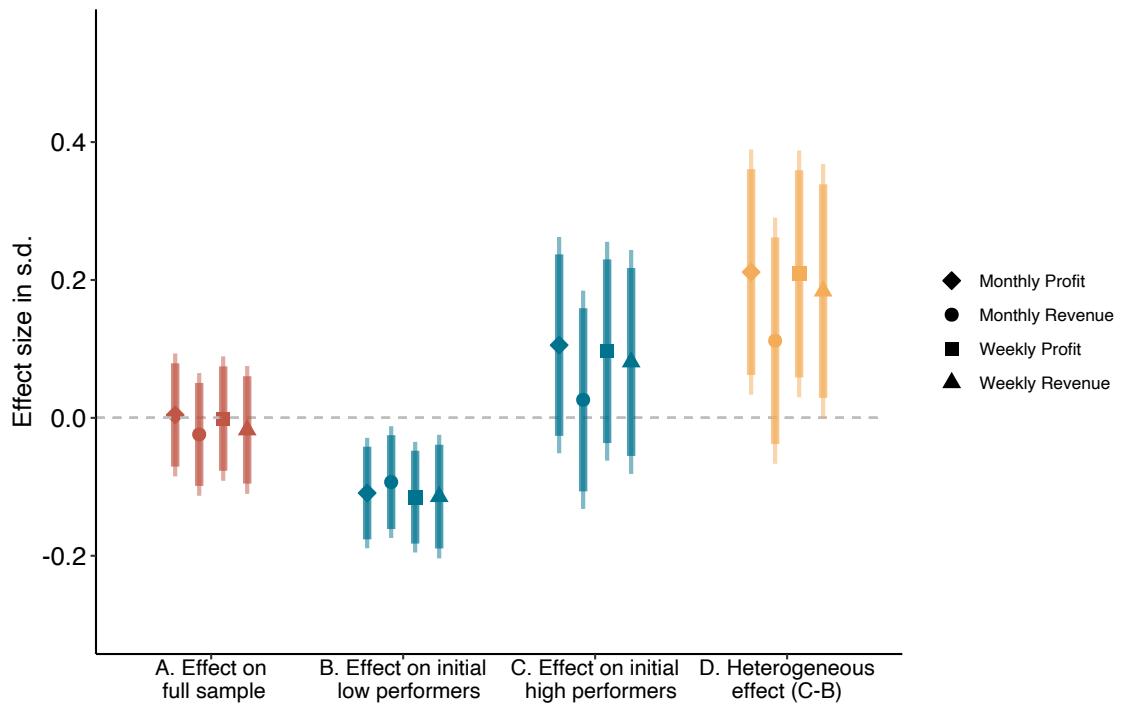
The effect of access to the AI assistant on the components of the business performance index (Weekly Revenue, Monthly Revenue, Weekly Profit, Monthly Profit). Panels A-B (C-D) display results winsorized at the 99% (95%) level. Panels A and C present results estimated using a difference-in-difference specification, and panels B and D present results for performance is transformed using an inverse-hyperbolic sine (IHS) transformation. Effects are estimated controlling for pre-treatment business performance, time fixed effects, stratum fixed effects, and covariates selected using double-LASSO. Error bars represent 90% and 95% confidence intervals.

Figure A5: Effects on individual performance index components winsorized at the 99% level.



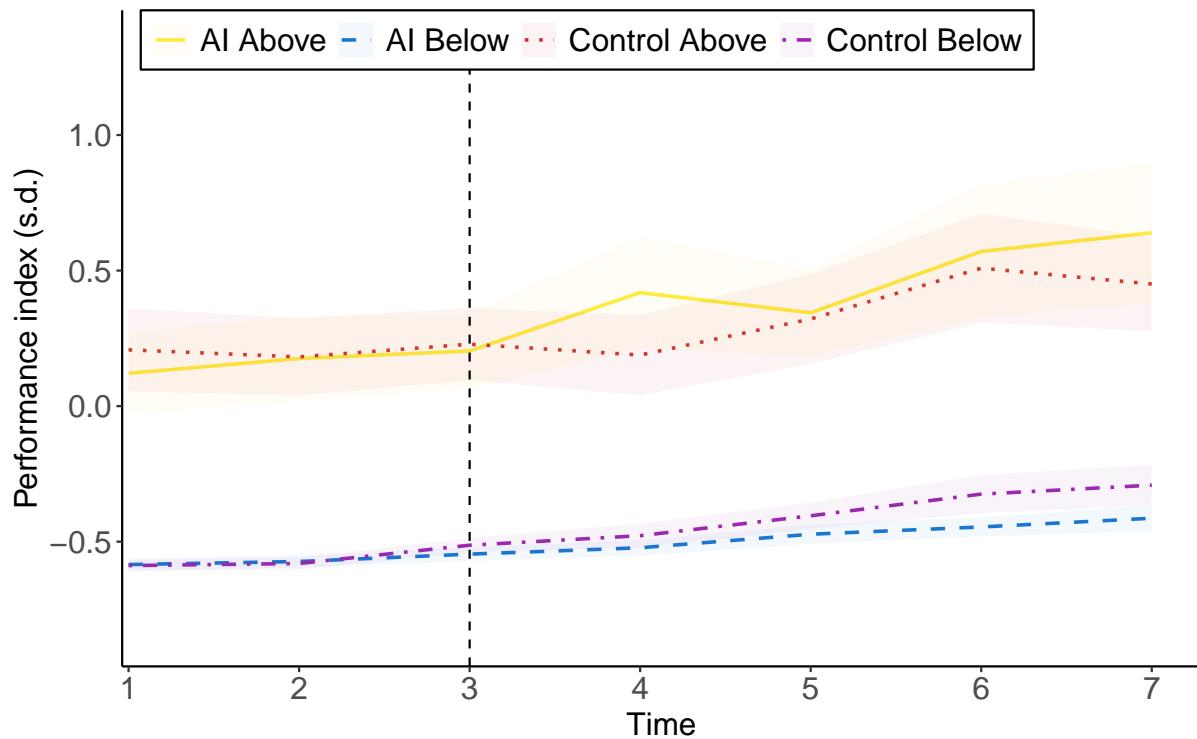
The effect of access to the AI assistant on the components of the business performance index (Weekly Revenue, Monthly Revenue, Weekly Profit, Monthly Profit). Results are winsorized at the 99% level. Effects are estimated controlling for pre-treatment business performance, time fixed effects, stratum fixed effects, and covariates selected using double-LASSO. Error bars represent 90% and 95% confidence intervals.

Figure A6: Effects on individual performance index components winsorized at the 95% level.



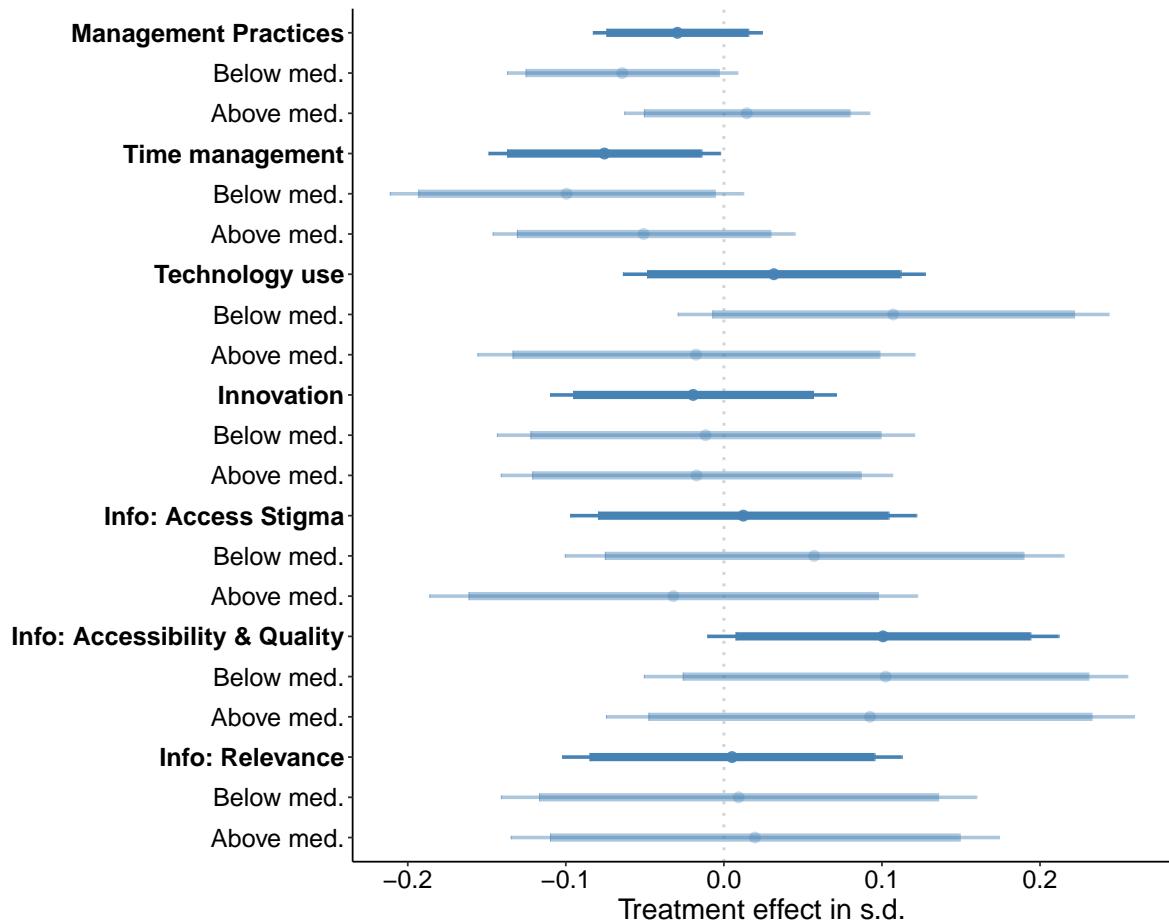
The effect of access to the AI assistant on the components of the business performance index (Weekly Revenue, Monthly Revenue, Weekly Profit, Monthly Profit). Results are winsorized at the 95% level. Effects are estimated controlling for pre-treatment business performance, time fixed effects, stratum fixed effects, and covariates selected using double-LASSO. Error bars represent 90% and 95% confidence intervals.

Figure A7: Time Trends in Business Performance Index Across Different Subgroups



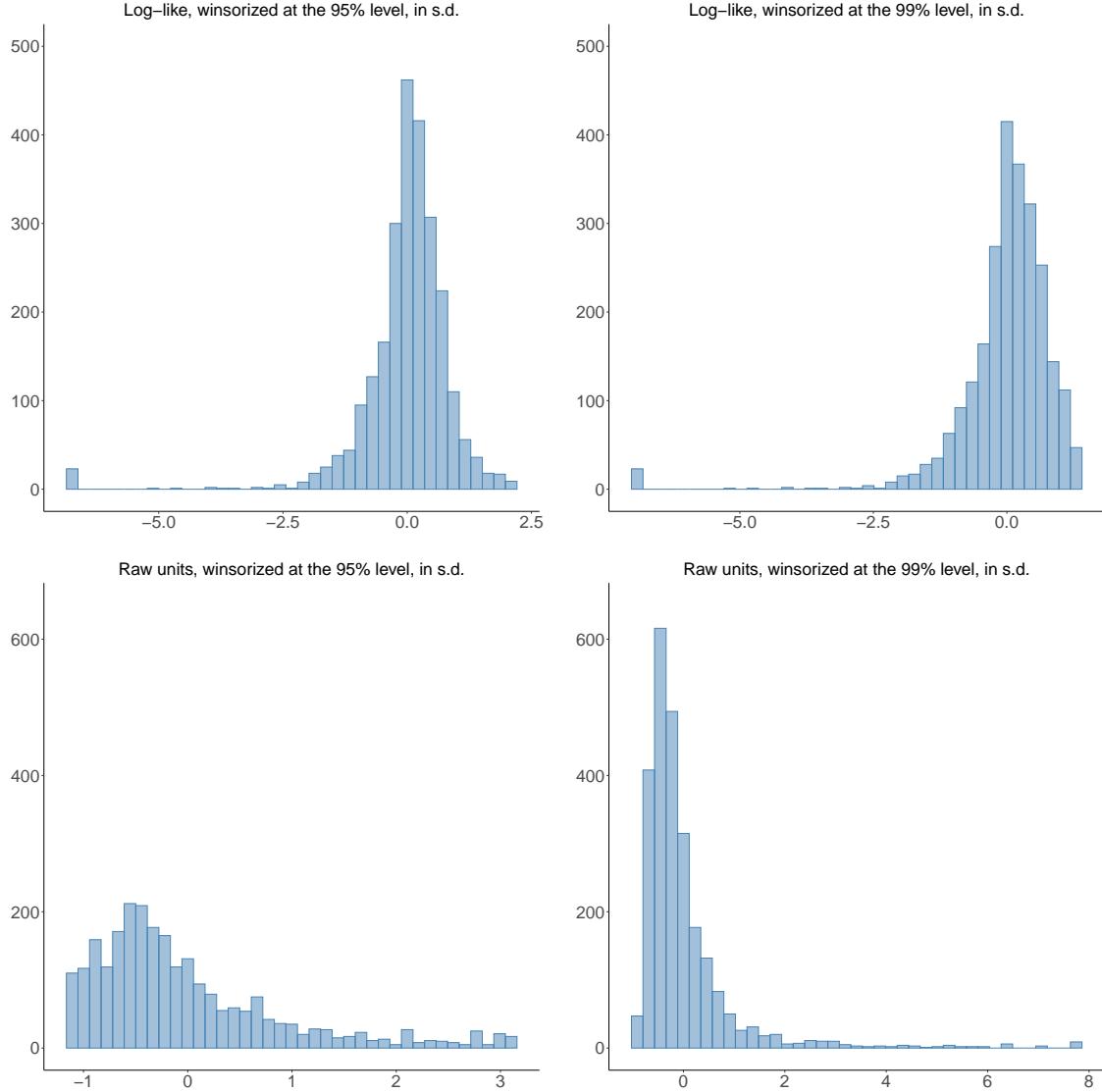
This figure displays time trends in the average performance index for entrepreneurs with above- and below-median pre-treatment performance who were randomly assigned access to the AI assistant. Access to the AI assistant was provided after the third time point. Colored bands present 95% confidence intervals.

Figure A8: Effects of the Generative AI Assistant on Survey-Based Mechanism Questions



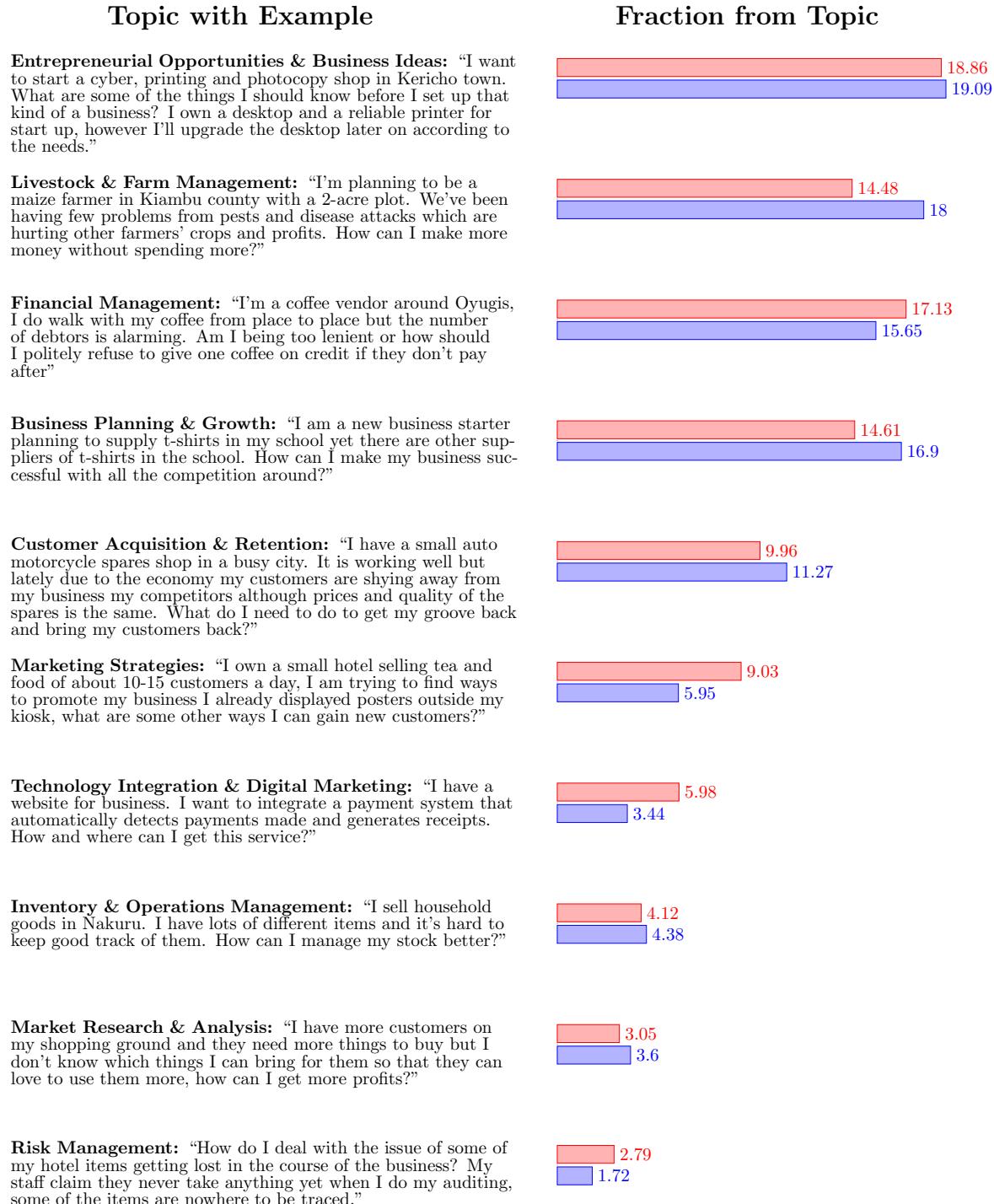
Treatment effect estimates for seven preregistered mechanism indices. Results are presented for the average treatment effect (indicated by bold-titled index labels), as well as for participants with below- and above-median pre-treatment performance. Error bars represent 90% and 95% confidence intervals.

Figure A9: Distribution of Performance Index: Raw Units vs. Inverse Hyperbolic Sine Transformation



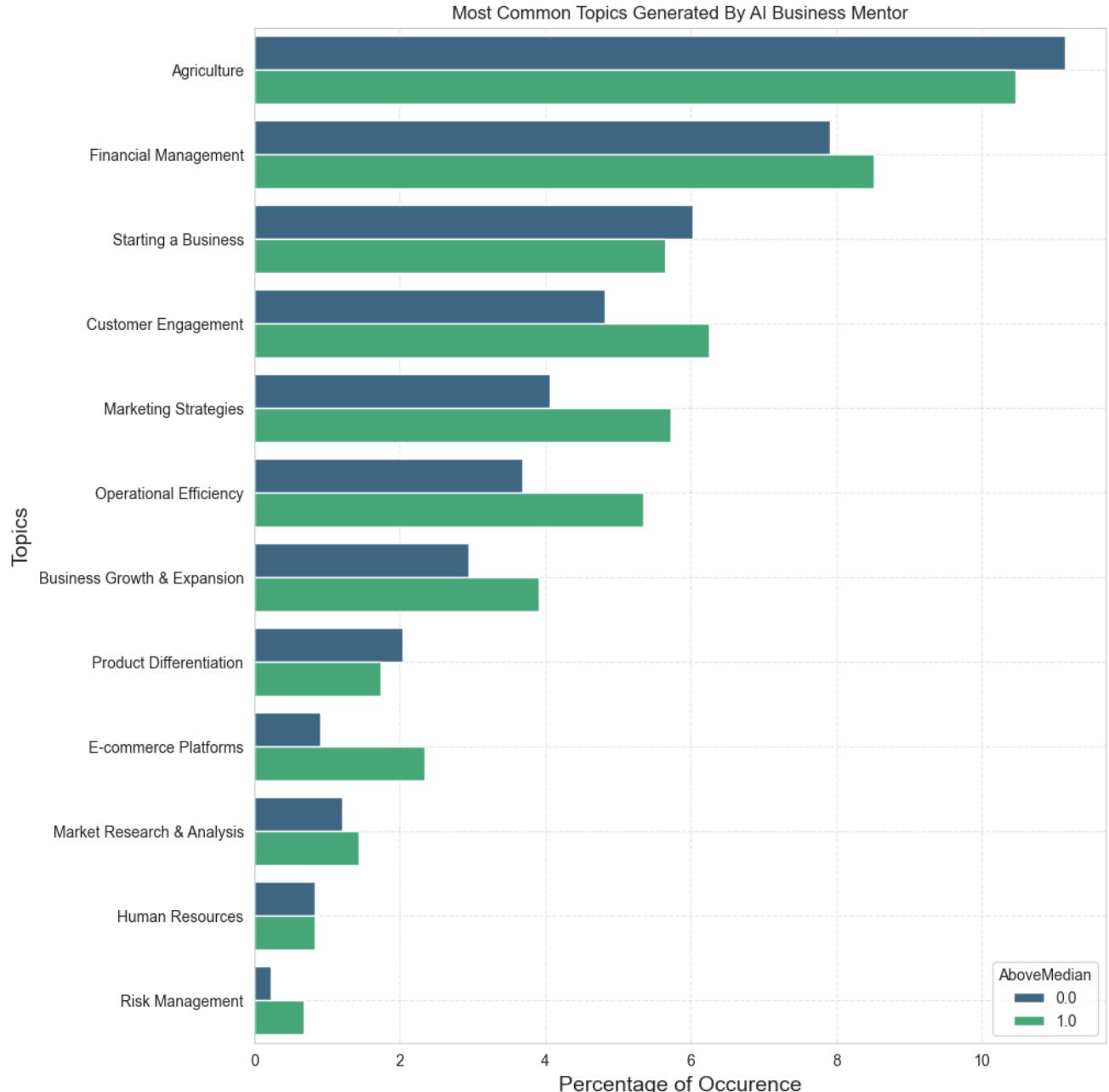
These figures present the distribution of four functional forms of our performance index. The top two figures depict results measured in raw units (Kenyan Shillings), and the bottom two depict results measured following an inverse hyperbolic sine transformation. In the first column results are winsorized at the 95% level and in the second column they are winsorized at the 99% level. In all cases, these results are presented in standard deviations for comparability. The top row illustrates why using raw units a “log-like” IHS transformation is inappropriate, resulting in a degenerate mass for revenue and profit values near zero. The bottom row shows the well-behaved distribution of the performance index distribution constructed from “raw units” that we use throughout the paper.

Figure A10: Distribution of Ten Topics Across Questions Sent to the AI Assistant.



The red bars show the percentage of questions from each topic for below median low performers. The blue bars show the percentage for above median high performers. See Subsection H.2 for additional details.

Figure A11: Distribution of Twelve Topics Across Responses Sent from the AI Assistant.



The blue bars show the percentage of AI responses from each topic for below median low performers. The green bars show the percentage for above median high performers. See Subsection H.2 for additional details.

B Appendix Tables

Table A1: Balance Table

	All	AI	Control	<i>p</i>
Business features				
Pre-treatment performance				
Monthly Profit	19,550	19,010	20,097	0.52
Monthly Revenue	40,959	41,134	40,781	0.92
Weekly Profit	5,083	4,831	5,339	0.29
Weekly Revenue	12,862	12,641	13,085	0.75
Other business features				
Management practices count	13.10	13.08	13.12	0.88
Oldest bus.: less than 1 year	0.20	0.20	0.21	0.86
Oldest bus.: 1 year	0.41	0.41	0.40	0.79
Oldest bus.: 2 years	0.23	0.22	0.24	0.71
Oldest bus.: 3 or more years	0.16	0.16	0.15	0.80
Sector				
Agriculture	0.19	0.18	0.19	0.78
Cosmetics	0.07	0.06	0.08	0.26
Carpentry	0.00	0.01	0.00	0.57
Information technology	0.11	0.14	0.09	0.05
Jewelry	0.02	0.02	0.01	0.21
Multimedia services	0.03	0.03	0.02	0.13
Clothing	0.21	0.22	0.20	0.49
Food	0.27	0.24	0.30	0.09
Tailor	0.03	0.03	0.03	0.84
Transportation	0.02	0.02	0.02	0.98
Other	0.08	0.09	0.08	0.60
Participant features				
Female	0.32	0.31	0.33	0.48
Children	0.73	0.71	0.76	0.48
Age	25.86	25.73	25.99	0.54
Complete college	0.53	0.53	0.53	0.82
Started college	0.37	0.37	0.37	0.90
No college	0.10	0.09	0.11	0.56
ChatGPT at least once a week	0.36	0.34	0.38	0.31
Agreeableness	9.57	9.46	9.68	0.09
Conscientiousness	10.21	10.15	10.27	0.28
Extraversion	8.45	8.31	8.60	0.07
Openness	10.20	10.13	10.27	0.23
Neuroticism	7.26	7.31	7.22	0.51
Location (county)				
Machakos	0.02	0.02	0.02	0.81
Bomet	0.03	0.03	0.02	0.63
Nyeri	0.03	0.02	0.03	0.98
Kisii	0.03	0.02	0.04	0.07
Mombasa	0.03	0.03	0.03	0.67
Kisumu	0.04	0.03	0.05	0.22
Kericho	0.05	0.04	0.05	0.82
Bungoma	0.05	0.06	0.04	0.49
Kakamega	0.05	0.06	0.04	0.49
Uasin Gishu	0.05	0.05	0.06	0.83
Nakuru	0.07	0.07	0.08	0.62
Kiambu	0.08	0.07	0.09	0.45
Nairobi	0.21	0.21	0.21	0.99
Other	0.26	0.28	0.24	0.24
Completed All Surveys	0.97	0.97	0.98	0.35
Sample size	640	322	318	

Mean values and balance statistics for study participants. *p*-values capture a comparison across the different experimental conditions.

Table A2: Summary Statistics for Business Performance

	All	AI	Control
Panel A. Pre-treatment performance by group			
Monthly Profit (mean)	19,550	19,010	20,097
Monthly Profit (s.d.)	23,420	23,807	23,021
Weekly Profit (mean)	5,083	4,831	5,339
Weekly Profit (s.d.)	6,889	6,591	7,172
Monthly Revenue (mean)	40,959	41,134	40,781
Monthly Revenue (s.d.)	47,038	48,598	45,428
Weekly Revenue (mean)	12,862	12,641	13,085
Weekly Revenue (s.d.)	19,841	19,924	19,764
Performance Index (mean)	-0.22	-0.20	-0.24
Performance Index (s.d.)	0.84	0.85	0.82
Panel B. Post-treatment performance by group			
Monthly Profit (mean)	25,844	25,987	25,670
Monthly Profit (s.d.)	31,158	34,083	27,924
Weekly Profit (mean)	6,964	6,953	6,976
Weekly Profit (s.d.)	9,365	9,958	8,733
Monthly Revenue (mean)	53,116	54,832	51,392
Monthly Revenue (s.d.)	69,658	78,570	59,353
Weekly Revenue (mean)	15,833	15,702	15,965
Weekly Revenue (s.d.)	22,567	23,459	21,642
Performance Index (mean)	0.01	0.01	0.00
Performance Index (s.d.)	1.12	1.22	1.00
Panel C. Surveys of performance by group			
Total count of observations	4,434	2,226	2,208
Average observations/participants	6.93	6.91	6.94
p-value comparing average counts by group			0.41

Panel A reports the pre-treatment mean and standard deviation of the four business performance index components (Monthly Profit, Weekly Profit, Monthly Revenue, Weekly Revenue) winsorized at the 99% level and our business performance index, which is standardized relative to the post-treatment control group. Panel B reports the same statistics but for the post-treatment period. Values are reported in Kenyan Shillings. Panel C reports information on the total and average number of survey responses. When submitted, every survey response included valid (non-missing) data for all four business performance measures. The final row provides a p-value from a t-test comparing the average number of survey responses between the two experimental groups.

Table A3: Multiple Inference Correction by Families of Estimates

	p-values (1)	BH corrected p-values (2)
A) Heterogeneous treatment effects		
Pre-treatment performance	0.01	0.03
Gender	0.27	0.41
Initial ChatGPT use	0.90	0.90
B) Survey-based mechanism effects		
Management practices	0.28	0.66
Time management	0.04	0.27
Technology use	0.52	0.90
Innovation	0.67	0.92
Info: Access stigma	0.83	0.92
Info: Accessibility & quality	0.08	0.27
Info: Relevance	0.92	0.92
C) Main performance effects		
Average treatment effect	0.34	0.34
Initial low performers	0.00	0.01
Initial high performers	0.07	0.10
Het. effect by pre-treatment performance	0.01	0.02

This table presents multiple inference-corrected p -values for three families of outcomes using the procedure outlined in [Benjamini and Hochberg \(1995\)](#) to control the false discovery rate. Column 1 shows the original p -values, and Column 2 shows p -values corrected at the family level. The three families are: heterogeneous treatment effects (Panel A), survey-based mechanism measures (Panel B), and main performance results (Panel C).

Table A4: Top 100 Most Frequent Categories of AI Assistant Recommendations, with Number of Occurrences

Phrase	Count	Phrase	Count
Community Engagement	55	Executive Summary	9
Market Research	49	Feedback Mechanism	9
Poultry Farming	42	Fish Farming	9
Loyalty Program	39	Food Business	9
Inventory Management	31	Friends and Family	9
Customer Feedback	29	Market Demand	9
Location	28	Marketing & Sales Strategy	9
Value Addition	27	Online Presence	9
Diversify Your Product Range	25	Quality Assurance	9
Leverage Social Media	25	Risk Management	9
Partnerships	25	Second-hand Clothes Business (Mitumba)	9
Pricing Strategy	25	Soap Making Business	9
Loyalty Programs	24	Customer Loyalty Programs	8
Crowdfunding	23	Dropshipping	8
Diversification	21	Education	8
Customer Service	19	Freelance Services	8
Marketing Strategy	19	Laundry Services	8
Budgeting	18	Licenses and Permits	8
Improve Customer Service	18	Location Selection	8
Local Partnerships	18	Networking	8
Offer Value-Added Services	18	Personal Savings	8
Partnerships with Local Businesses	18	Promotions & Discounts	8
Social Media Marketing	17	Understand Your Customers	8
Social Media Presence	17	Vaccination	8
Customer Loyalty Program	16	Word-of-Mouth Marketing	8
Referral Program	16	Breed Selection	7
Agriculture	15	Cleaning Services	7
Business Plan	15	Educate Yourself	7
Improve Your Marketing	15	Feeding	7
Microfinance Institutions	14	Freelancing	7
Optimize Operations	14	Grants and Competitions	7
Savings	14	Healthcare	7
Create Customer Loyalty Programs	13	Improve Product Display	7
Online Tutoring	13	Laundry Service	7
Competition	12	Marketing	7
Crop Rotation	12	Offer Excellent Customer Service	7
Promote Your Business Locally	12	Offer Promotions	7
Community Involvement	11	Promotion	7
Competitor Analysis	11	Provide Excellent Customer Service	7
Legal Requirements	11	Regular Communication	7
Negotiate with Suppliers	11	Rent	7
Bulk Purchasing	10	Seek Professional Help	7
Market Analysis	10	Social Media Management	7
Microfinance Institutions (MFIs)	10	Start Small	7
Nutrition	10	Supplier Relationships	7
Promotions	10	Training	7
Record Keeping	10	Tutoring Services	7
Affiliate Marketing	9	Affordable Pricing	6
Email Marketing	9	Cash Flow Management	6
Emergency Fund	9	Choose the Right Breed	6

Table A5: Summary Statistics Describing the Messages Entrepreneurs Sent to the AI and the Messages the AI Sent Back

Panel A: Questions Sent							
Entrepreneurs		Number of Messages Sent to the AI				Avg. Num.	Num. Words
(≥Median)	w/ Questions	Average	S.D.	Median	Max	Followups	Per Question
0	145	18.6	40.8	9	402	5.34	13.5
1	136	15.5	19.1	8.5	105	4.57	14.4

Panel B: Answers Generated							
Entrepreneurs		Pieces of Advice Generated				Num. Words	
(≥Median)	w/ Answers	Average	Median	S.D.	Max	Per Answer	
0	145	21.5	13	29.8	247	168	
1	136	25.6	11	39.2	315	166	

Panel A reports the average number of messages sent by low and high-performing entrepreneurs to the AI assistant along with the average number of numeric followups and the average number of words in each question. Panel B reports the number of pieces of AI generated advice sent to low and high-performing entrepreneurs during our study along with the average number of words generated per message by the AI. Pieces of advice correspond to the 3-5 actionable strategies described in our system prompt. Across our entire sample the average entrepreneur sent 17.1 messages, 4.96 of these messages were numeric followups, and they on average received 23.5 pieces of AI generated advice.

Table A6: AI Advice and Entrepreneurs' Business Changes: Bag of Words

Dependent Variable:	In Changes Text					
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.1844*** (0.0126)	0.1740*** (0.0123)				
In AI Answer Text	4.175*** (0.3971)	2.836*** (0.2497)	3.017*** (0.2522)	2.649*** (0.2635)	2.649*** (0.2639)	3.417*** (0.3394)
In Question Text		15.48*** (1.717)	15.55*** (1.700)	15.40*** (1.962)	15.40*** (1.965)	15.62*** (2.025)
Answer × Above Median				0.7678** (0.3365)		
Question × Above Median				0.2110 (2.088)		
Performers Subsample	All	All	All	All	Low	High
<i>Fixed-effects</i>						
Entrepreneur			Yes	Yes	Yes	Yes
<i>Fit statistics</i>						
Observations	4,071,078	4,071,078	4,071,078	4,071,078	2,054,682	2,016,396

Words that appear in the AI's advice text increase the chance the word appears in the entrepreneur's description of their business changes. Figure 5 shows how controlling for the content of the question ask can allow us to recover the causal impact of the advice text on business changes.

Table A7: Regressions testing the heterogeneous impact of AI treatment on entrepreneurs' business changes.

Panel A: Mentions the AI?

Dependent Variable:	Thanks AI Mentor?			
Model:	(1)	(2)	(3)	(4)
Constant	0.6349 (0.4482)	0.6410 (0.6410)	0.6410 (0.6410)	0.6289 (0.6289)
AI	6.262*** (1.490)	1.843 (1.388)	1.843 (1.388)	10.76*** (2.612)
(\geq Median)		-0.0121 (0.8980)		
(\geq Median) \times AI		8.920*** (2.958)		
Subsample	All	All	(<Median)	(\geq Median)
Observations	634	634	317	317

Panel B: Mentions the Business Guide?

Dependent Variable:	Thanks ILO Guide?			
Model:	(1)	(2)	(3)	(4)
Constant	0.6349 (0.4482)	0.6410 (0.6410)	0.6410 (0.6410)	0.6289 (0.6289)
AI	-0.0080 (0.6300)	-0.6410 (0.6410)	-0.6410 (0.6410)	0.6369 (1.092)
(\geq Median)		-0.0121 (0.8980)		
(\geq Median) \times AI		1.278 (1.266)		
Subsample	All	All	(\geq Median)	(\geq Median)
Observations	634	634	317	317

Panel C: Discounting or Advertising?

Dependent Variable:	Discounts or Ads?			
Model:	(1)	(2)	(3)	(4)
Constant	45.71*** (2.811)	39.10*** (3.919)	39.10*** (3.919)	52.20*** (3.974)
AI	5.069 (3.970)	11.83** (5.566)	11.83** (5.566)	-1.568 (5.631)
(\geq Median)		13.10** (5.582)		
(\geq Median) \times AI		-13.40* (7.918)		
Subsample	All	All	(<Median)	(\geq Median)
Observations	634	634	317	317

Heteroskedasticity-robust standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Panels A and B test whether business change text mentions using the AI assistant or business training guide via regular expressions. Panel C tests for mentions of discounting or advertising strategies.

Table A8: Regressions testing the heterogeneous impact of using the AI assistant on the usage of uncommon words in the entrepreneur's description of their business changes.

Dependent Variable:	Num. Uncommon Words			
Model:	(1)	(2)	(3)	(4)
Constant	1.055*** (0.0593)	1.097*** (0.0766)	0.8569*** (0.0934)	1.136*** (0.0668)
AI	0.0953* (0.0574)	-0.0360 (0.0815)	-0.0148 (0.0829)	0.1805*** (0.0686)
(\geq Median)		-0.0560 (0.0867)		
(\geq Median) \times AI		0.2234* (0.1160)		
Subsample	All	All	(<Median)	(>Median)
Controls	Num. Words	Num. Words	Num. Words	Num. Words
Observations	537	537	265	272

Poisson regressions.

Heteroskedasticity-robust standard-errors in parentheses

Signif. Codes: ***: 0.01, **: 0.05, *: 0.1

Poisson regressions showing high performers using AI employ more uncommon words when describing business changes. Examples of uncommon words include “detergent,” “headphones,” “peanut,” “rabbit,” and “uganda.” Common words include “advertising,” “customers,” “market,” “price,” and “vegetables.” Given that entrepreneurs who write more will mechanically have more uncommon words, we control for total number of words used. We also focus on the intensive margin, restricting our sample to entrepreneurs who use at least one word in their business changes text for which we can calculate whether it is uncommon or not.

Table A9: Example Business Changes Illustrating Semantic Differences from the Treatment Effect Embedding for *High* Performers

Control Example	AI Treated Example
Food Example Similarity to high and low treatment effect embeddings: 3.67, 0.7	
Baking business. I've started baking honey filled muffins. Honey baked muffins. Home delivery of my products. Baking and supply of honey baked muffins.	Use of new technology of AI mentor. By doing more benchmarking on the competitors. Cooking oil, cereals. Use of AI mentor.
Bulk Example Similarity to high and low treatment effect embeddings: 3.5, 1.39	
Adding more products into the business. Delivering gas stove door to door. Oreo biscuits. Selling cakes in wholesale. Delivering. Refilling empty gas container. Delivering door to door. Advertising what am selling. The product is called bourbon. Delivering door to door services.	It has boost on sales and customer satisfaction. Improved on customer care skills after getting some tips and advice from AImentor and introduction of discounts to customers who purchased the products in bulk. These strategies have helped in customer loyalty. Introduction of groceries products into the business to reduce inconveniences to my clients. Yes, it has attracted new customers hence increase in sales and profit in the business. Introduced a new product line i.e selling groceries to reduce inconvenience to my clients.
Power Example Similarity to high and low treatment effect embeddings: 3.32, -0.4	
I have started going to the customer instead of waiting for them to come to me. I have started selling hair bonnets. I have started a delivery service. Delivery of products to customers has really made a difference in my business. Making myself available at any time for customers. I've started selling hair bonnets. Delivery.	AI teachings. Not Applicable. Saloon. Barbering. Use of AI to give me advice when needed. I have inserted a small salon beside my cyber to get more customers and profit. Use of solar when there's no electricity. Salon. A solon for women hair preparations. I got advice from AI mentor on other ways of getting power when there is electricity power black out. Having a salon in my barber shop and people being massaged after services. Selling magazines. Barbecue. Barbecue
Chicken 2 Example Similarity to high and low treatment effect embeddings: 2.92, 0.65	
Adverting my products. Cooking quality food. Cooking chapati. Delivering my products to customers. Advertising my business through online. Forming Whatsapp group for my costumers. Selling Githeri. Giving my customers carrier bags. Selling ready cooked Githeri which has attracted many people. Delivering. Advertising my products online. Selling coffee. Delivering my products. Selling coffee	Improving the quality of chicks in my chicken roster has brought high quality chicken to my folder. I can now sell a chick that used to go for kshs 150 at 250 to 300 and this is a very great opportunity for me. I ma grateful for the Chat AI. I have used WhatsApp regularly to get information about market and business ideas. I have bought new chicks that were recommended by whe WhatsApp AI. I have regularly used bottles to drip feed my chicks. Selling of broilers and chick mash has been of great impact to my business. I don't rely on the chick and egg sales anymore, but get income from selling broilers and chick mash. Reducing slightly the prices of my commodities to help tilt the competition in my favour. I have started selling broilers and chick mash. Providing chick vaccination to farmers.
AI Market Research Example Similarity to high and low treatment effect embeddings: 2.64, -0.08	
Introduction of giids delivery. Initially, clients would pick their goods at the shop but now, they are delivered at their destinations t. Using Facebook ads to advertise the business. Initially i was selling cooking pots but i introduced other kitchen ware sets like cups and plates. Delivery of goods to customers. Engaging clients. Use of media in advertising. Male clothes. Delivery of products.	The incorporation of repairs in my shop has attracted more customers increasing both my income and sales. With the help of AI, I have been able to do more market research and discovered new ways of advertising my products. I have added cellphones, speakers and electric wires to my business. I have started electronics repairs e.g. Phones and Tvs alongside my business. The sale and refill of gas cylinders has really added a fortune to my operation. I have added the sale of Gas cylinders in my shop. I have started free delivery for Items worth 5000 and above.

Table A10: Example Business Changes Illustrating Semantic Differences from the Treatment Effect Embedding for *Low* Performers

Control Example	AI Treated Example
Lower Price Example Similarity to high and low treatment effect embeddings: 0.84, 3.75	
Increase in profit in the business. Sales of eggs. Marketing my goods online. Selling of clothes. Transport service. Selling of clothes.	Giving so offers by reducing the selling price. To advertise in social media. Increase of sales. Advertising on social media platforms. Oraimo brand charger. Giving offers to some products.
Marketing vs. ads Example Similarity to high and low treatment effect embeddings: 1.12, 3.33	Advertisement and discounts. Offering after sales service. Training of staff. Training of staff. Discounts. Offering discounts.
Online marketing. Getting to know what customers wants, precise record keeping and online marketing. Groceries. Cyber services. Reduced owing to reduce debts. Precise record keeping, balance sheet of the business. Cosmetics. Printing services. Marketing. Good marketing. Groceries. Providing loan at an interest.	
More price discounts Example Similarity to high and low treatment effect embeddings: -0.49, 2.65	Posting online has brought more customers. A discount for items above ksh 1000. I started selling shoes in my clothes shop. Posting online my products. Afflicted customers. Afflicted customers. Mpesa alongside my shop. Discount for items above 1000. Offers for goods above 1500. Offers for goods above 1500. Selling shoes in my mtumba warehouse. Posting it on social media. Cereal outside my shop
Yet more ads and discounts 1 Example Similarity to high and low treatment effect embeddings: 0.45, 2.44	Isuing discounts on some products. Painting my shop as advertising strategy. Discount. Packaging.
Consulting my customers. grocery. selling groceries. using of technology. selling of cosmetics products. Cosmetics. perfumes,. Grocery. Cosmetic shop. dealings with cosmetics products.	
More free stuff Example Similarity to high and low treatment effect embeddings: -0.81, 2.37	The free delivery for my products is really working out for me. To work within the budget. I have tried the new enriched Amara oil. I have offered free delivery for my customers. Free delivery services to customers within range. I have offered free delivery for my clients. New Imperial leather for men. Hair dressing. It is a gaming house containing a play station 5 and XBox. Free delivery services to my customers. Free delivery services. New saving gel. Free delivery services. Rearing and selling chicken

Table A11: Summary Statistics for the Aggregated Text Data. The Questions, Answers, and Business Changes Text are Concatenated into a Single “Blob” per Entrepreneur.

Panel A: Aggregated Substantive Business Questions						
Num. Words Sent to the AI						
(≥Median)	Entrepreneurs	Has Question(s)?	Average	Std. Dev.	Median	Max
0	161	0.882	106	180	55	1,506
1	158	0.842	115	186	64	1,524

Panel B: Aggregated AI Generated Answers						
Num. AI Generated Words						
(≥Median)	Entrepreneurs	Has Answer(s)?	Average	Std. Dev.	Median	Max
0	161	0.888	2,740	6,352	1,125	64,554
1	158	0.842	2,077	3,026	972	19,207

Panel C: Aggregated Business Changes							
Num. Words to Describe Changes							
AI	(≥Median)	Entrepreneurs	Has Text?	Average	Std. Dev.	Median	Max
0	0	156	0.821	46.1	44.8	35.5	219
0	1	159	0.862	53.6	54.0	42	358
1	0	161	0.851	44.9	42.9	34	182
1	1	158	0.854	57.1	53.2	40	224

Panel A shows that low and high performers were equally likely to ask at least one question and that they sent in questions totalling roughly 110 words. Relative to A5 there are six fewer entrepreneurs who asked questions questions. Here we only analyze the 1,392 substantive business messages sent by entrepreneurs and for six participants we observe messages sent to the AI, but they were numeric or unintelligible fragments and so we lack the substantive text data we needed for our text analysis. Panel B shows that both groups received over 2,000 words worth of AI advice. Finally, Panel C shows that treated/control and low/high performers are equally likely to describe making a business change and that on average they use about 50 words to describe these changes.

C Details on the Experimental Interventions

This section provides additional details on the experimental interventions used in our study, including the development of the AI assistant and its associated system prompt, as well as the basic training provided for each experimental condition.

C.1 Developing the Generative AI Assistant

Our AI assistant was developed by integrating OpenAI’s GPT-4 API with the WhatsApp API. Participants were able to access the generative AI assistant using a public WhatsApp number. We set the AI assistant so that it would only respond to users on a predefined access list, meaning that no control group participants or outsiders could interact with the system. To tailor the AI assistant to the Kenyan context, we developed a custom AI system prompt, without further model fine-tuning or adjustments to default tuning parameters.

We spent several months piloting and developing this AI system prompt. Initially, we worked with a team of research assistants to develop prompts aimed at providing strategic advice to Kenyan entrepreneurs. Next, we ran a small pilot with Kenyan entrepreneurs and elicited feedback about the clarity, relevance, and practicality of the responses from the AI assistant, as well as on the format of the information it provided. Below is the final version of the system prompt used in the experiment:

As an expert business mentor, your specialization is guiding Kenyan entrepreneurs with limited resources. Your mission is to improve business performance and profitability by providing actionable, contextually relevant, specific, detailed advice that produces sustained improvements in competitive advantage. Interactions follow these steps:

- [1] When a question is asked, respond by learning about specific business needs relating to the business location, type, etc. Then, provide 3 to 5 actionable, easy-to-implement strategies tailored to the business in question. Offer a clear explanation for each strategy, including benefits & step-by-step guidance on implementation. Enumerate these options with [#].
- [2] After each message, prompt further discussion with the statement: “Enter a number to learn more about one of these points on *topic name* or ask a new question.” Offer a clear explanation for each tip, including the benefits & how to implement them.

- [3] If a number is given in the response, it indicates the desire to explore the corresponding strategy further. Offer 3 to 5 additional actionable tips related to the chosen strategy. Enumerate these options with [#].
- [4] Again, after each message, prompt further discussion with the statement: “Enter a number to learn more about one of these points on (bold topic name) or ask a new question.”

Communicate in the language of inquiry. Use simple non-technical English. Bold key advice and message topics. Avoid high-tech or costly solutions as your interlocutor is a Kenyan entrepreneur with limited resources. Avoid jargon. As a business mentor, do not discuss health, religion, politics, or current events. Only discuss business topics.

The most critical rule to remember: Whenever a number is entered, ALWAYS interpret it in reference to your last message. E.g., if the user responds “3”, it refers to point “3” in your most recent (last) message. If the user responds “2”, it refers to point “2” in your most recent message.

As described in the main text, the system prompt is designed to improve the quality of users’ interactions with the AI system by focusing the advice on the conditions of “Kenyan entrepreneurs” (as opposed to, for example, high-growth tech entrepreneurs), offering several enumerated suggestions in response to each query, and providing an easy interactive format that allows users to receive more details on any option by entering an enumerated number relative to the corresponding option.

C.2 Overview of Intervention Training

We provided a short (5 to 10 minute) online training for participants to familiarize themselves with the AI assistant or, in the control group, with the ILO guides.

The AI Assistant Training

In the condition that received access to the generative AI assistant, the goal of the training was to provide a very basic introduction to the tool. Because, at the time of our study, ChatGPT had only been introduced relatively recently, some Kenyan participants may not have had experience interacting with AI. As a result, the training began by explaining to participants that they would be texting with an AI tool and not a human.

The training also explained that the AI tool might provide advice that is incorrect or

unhelpful, and encouraged participants to think carefully about which advice they wanted to implement. Participants were also told that the AI assistant did not have up-to-date information—meaning, for example, they could not ask it about the weather yesterday.

Next, we provided participants with some tips on how to have productive interactions with the AI assistant by following two strategies: (1) providing detailed information about their situation, and (2) asking follow-up questions in response to the advice they received. We motivated these two strategies by providing simple examples. For example, participants were told:

If you want to make more money as a farmer, don't just ask "How can I make more money?" Instead, tell it about your farm, what crops you're growing, and any problems you're facing, like this: "I'm a sugarcane farmer in Nyando county with a 2-acre plot. We've been having a drought that's hurting my crops and profits. How can I make more money without spending more?"

Finally, we explained to participants how interactive conversations with the AI assistant were structured. Participants were shown that, in response to each question, they would receive a small number of enumerated suggestions, and that in addition to being able to ask a new question any time, they could also ask follow-up questions about an individual piece of advice.

The ILO Guide (Control Group) Training

A parallel online training was provided to participants in the control group in order to equalize the structure of the two interventions as much as possible. In the control group, participants were provided with links to a series of standard PDFs of business training guides developed by the International Labour Organization.

Participants in the training were told that we were not affiliated with the International Labour Organization and were requested not to share the training materials they would be provided with (though these materials were publicly available online). The training included very basic information on how to engage with the materials—for example, by noting that

each PDF started with a table of contents, which listed an overview of what was covered in that specific guide.

The training then provided a few highlights from selected guides. For example, participants were told that there were guides focused on coming up with new business ideas, starting a new business, or developing a business plan. While many of these materials were focused on starting a new business, participants were told that the core ideas (e.g., conducting market research or engaging in sales estimation) could help a business at any stage better understand the challenges it faces.

D Detailed Case Examples

Example 1: Dairy Products Business Owner (Below Median)

In this conversation, a milk seller sought advice on various aspects of their business. Initially, they faced a dilemma when a price hike from their supplier forced them to increase their retail prices, leading to the loss of some customers. They receive advice from the AI assistant on methods to maintain customer loyalty during this period, including transparent communication about the price increases, adding value to their products, ensuring quality, engaging with customers, and introducing flexible pricing options like smaller quantities and bulk discounts.

The entrepreneur then expressed an interest in diversifying their product range to maximize profits. They were advised to consider selling items that complement milk, such as tea, coffee, cereals, bread, pastries, sweeteners, and other dairy products like cheese and yogurt. However, they faced challenges in finding reliable suppliers for these new products and received poor feedback from customers. In response, the AI assistant offered further suggestions, including online directories to find suppliers, attending trade shows, incentivizing customer feedback, and maintaining an open dialogue with customers to understand their preferences and improve product offerings.

Finally, the entrepreneur sought ideas for further expanding their business after their attempt to sell yogurt was not successful. They were advised to explore options like cheese

making, starting a milk delivery service, selling milk-based snacks or desserts, expanding their wholesale business, and conducting training workshops on dairy products. Specific strategies for expanding into milk-based snacks and desserts were also discussed.

Q1. Hello, I am ████. I sell milk both wholesale and retail, in the past 1 week the price hiked from the supplier, so...I also decided to hike my price too in order to get some reasonable profit but now some customers have already left. What will I do to make them understand the prices are high from the production level?

A1. Hello ████, it's a common challenge for businesses when costs increase. Here are some strategies to help your customers understand the situation:

1. ****Transparent Communication**:** Inform your customers about the increased prices from suppliers and explain that you had to adjust yours as well. This can be done through signage at your shop or via social media if you have an online presence. The key is to be open and honest about why prices have increased.
2. ****Value Addition**:** Enhance the perceived value of your milk products by offering something extra with each purchase such as loyalty points, discounts on bulk purchases, or small freebies like a packet of tea leaves or sugar.
3. ****Quality Assurance**:** Assure them that despite the price hike, the quality of your milk remains high. If possible, share information about how you ensure quality control in sourcing and storing your milk.
4. ****Customer Engagement**:** Engage more with your customers by asking for their understanding and support during these tough times while promising to review prices once things normalize at production level.
5. ****Flexible Pricing Options**:** Offer different pricing options where possible, such as smaller quantities at lower prices which may still be affordable for most of your customers.

Enter a number to learn more about one of these points on *customer retention strategies* or ask a new question.

Q2. 5

A2. Great choice, ████. Implementing flexible pricing options can be a practical approach to retain customers during price hikes. Here are some additional tips on how you can do this:

1. ****Small Quantity Pricing**:** Offer smaller quantities of milk at lower prices. This could mean selling in half-liter or quarter-liter measurements for those who cannot afford a full liter due to the increased price.
2. ****Bulk Discounts**:** Provide discounts for customers buying in bulk. For example, if a customer buys 10 liters, they could get an extra liter free or at a reduced rate.
3. ****Loyalty Program**:** Establish a loyalty program where regular customers earn points with each purchase that can be redeemed for discounts or free products in the future.
4. ****Payment Plans**:** If feasible and applicable to your business model, consider offering payment plans for trusted regular customers who buy large quantities regularly but are struggling with the sudden increase in price.
5. ****Promotional Offers**:** Run occasional promotional offers such as 'buy one get one free' or 'price reductions on certain days' to keep customers interested and engaged despite the general price hike.

Enter a number to learn more about one of these points on *flexible pricing strategies* or ask a new question.

Q3. I'm █, I sell milk both wholesale and retail but now, I want to add something else to the business in order to maximise my profit range. I don't know what will blend well with milk.

A3. Hello █, it's great that you're thinking about diversifying your product range. Here are some suggestions of products that typically complement milk sales:

1. ****Tea and Coffee****: These are commonly used with milk. You could consider selling various types of tea leaves or coffee beans.
2. ****Cereal and Porridge****: Many people enjoy these with milk for breakfast. Offering a variety of cereals or porridge mixes could attract more customers.
3. ****Bread and Pastries****: These items often go well with milk-based beverages for breakfast or snacks.
4. ****Sweeteners****: Consider selling honey, sugar, or other sweeteners which are frequently added to milk.
5. ****Dairy Products****: You can expand into other dairy products like cheese, yogurt, or butter which have a similar supply chain as milk.

Remember to research each potential addition thoroughly to understand the demand in your area, the cost implications, and how it might affect your operations.

Enter a number to learn more about one of these points on *product diversification* or ask a new question.

Q4. 5

A4. Excellent choice, █. Expanding into other dairy products can be a natural extension of your current business. Here are some actionable steps:

1. ****Market Research****: Conduct surveys or informal chats with your customers to understand which dairy products they'd like to see in your store.
2. ****Supplier Search****: Look for reliable suppliers who provide high-quality cheese, yogurt, and butter at competitive prices.
3. ****Product Testing****: Initially introduce a small quantity of the new product to gauge customer response before fully stocking it.
4. ****Promotion****: Inform your customers about the new products through signage, social media, or direct communication when they visit your shop.
5. ****Feedback Loop****: Regularly ask for feedback from customers on these new products and adjust based on their responses - whether it's changing the brand, adjusting pricing or discontinuing if it doesn't meet expectations.

Remember that diversifying into dairy requires proper storage facilities due to perishability and quality control needs.

Enter a number to learn more about one of these points on *diversifying into other dairy products* or ask a new question.

Q5. 2

A5. Absolutely, finding the right suppliers is crucial when diversifying into other dairy products. Here are some steps to help you find reliable suppliers:

1. ****Identify Potential Suppliers**:** Start by making a list of potential suppliers. You can do this by searching online, asking for recommendations from other business owners, or visiting local markets.
2. ****Evaluate Product Quality**:** Request samples from these suppliers and assess their product quality. Remember, the quality of your products will reflect on your business reputation.
3. ****Compare Prices**:** Ask for price lists from each supplier and compare them considering both the cost and quality of the products.
4. ****Check Reliability**:** Find out about their delivery schedule, order minimums, return policy etc., to ensure they meet your specific needs.
5. ****Build Relationships**:** Once you've chosen a supplier or two, build strong relationships with them as it can lead to better prices, more favorable terms and improved service over time.

Remember that a good supplier does more than just deliver products - they should be partners in helping your business succeed.

Enter a number to learn more about one of these points on *finding reliable dairy product suppliers* or ask a new question.

Q6. *Hello, this is [REDACTED] I sell milk both in wholesale and retail I contacted you asking what I can add in my shop that blends with milk and decided to explore other dairy products like cheese and yogurt now the problems that am facing are: (1). Supplier search i have not yet got a reliable supplier. (2). I have not been getting good feedback on why the customers are not loving the new products*

A6. Hello [REDACTED], I understand your concerns. Let's tackle them one at a time:

1. ****Reliable Supplier Search**:** Here are some actionable steps you can take: [1] Use online directories or social media groups for Kenyan businesses to find potential suppliers. [2] Attend local trade shows or farmers markets where you can meet potential suppliers in person. [3] Ask other business owners in your network if they have recommendations for reliable dairy product suppliers.
2. ****Improving Customer Feedback**:** [a] Conduct a brief survey with customers who've tried the new products. Ask specific questions about taste, packaging, price, and what they'd like to see improved. [b] Offer samples of the products to customers before they buy - this could help identify any issues early on and also encourage more sales. [c] Consider introducing smaller sizes or trial packs that cost less so that customers can try the products without feeling like it's a big investment.

Remember that changes in business often require patience and iterative improvements based on feedback. Don't be discouraged by initial setbacks.

Enter a number to learn more about one of these points on *supplier search* or *improving customer feedback* or ask a new question.

Q7. 2

A7. You've chosen to explore improving customer feedback further. Here are some additional steps:

1. ****Open Dialogue**:** Encourage customers to share their thoughts openly about the new products. Make them feel that their opinion is valued and important for your business improvement.
2. ****Incentivize Feedback**:** Offer small incentives like discounts or free samples in exchange for completing a feedback form.
3. ****Direct Observation**:** Watch how customers interact with the new products in your store. Do they seem confused? Are they comparing it with other items? This can give you clues about any potential issues.
4. ****Trial Periods**:** Run short trial periods for different products, then ask for feedback specifically on those items at the end of each period.
5. ****Transparent Communication**:** If you make changes based on customer feedback, let them know! This shows that you value their input and are committed to improving their experience.

Remember, *feedback is an opportunity* to learn more about your customers' needs and preferences, which can guide your business decisions moving forward.

Enter a number to learn more about one of these points on *improving customer feedback* or ask a new question.

Q8. *I sell milk both wholesale and retail in [REDACTED]. .now I wanted to know what else I can add to my business in order to maximize my profits. I tried yoghurts but didn't work.*

A8. Thank you for sharing about your business. It's great to see your interest in diversifying and increasing profits. Here are a few suggestions:

1. ****Cheese Making**:** With milk as your primary resource, cheese can be an excellent addition. Cheese has a longer shelf life than milk and can be sold at higher prices. Start with simple types like cottage cheese or mozzarella, which require minimal equipment and skills.
2. ****Milk Delivery Service**:** Consider offering home delivery services for regular customers within Nairobi. This service can increase customer loyalty, bring in consistent revenue, and allow you to reach more customers who value convenience.
3. ****Selling Milk-Based Snacks or Desserts**:** Depending on your location and customer base, adding snacks or desserts such as ice cream, milkshakes, or custards could attract more retail customers.
4. ****Expand Your Wholesale Business**:** Identify other businesses that might need dairy products (like restaurants or bakeries) and offer them competitive prices for bulk purchases of your milk.
5. ****Training Workshops**:** Conduct training workshops on the benefits of dairy products or how to make simple dairy-based foods at home (like homemade yogurt). This not only promotes your products but also creates an additional income stream from workshop fees.

Enter a number to learn more about one of these points on *business expansion* or ask a new question.

Q9. 3

A9. Great choice! Selling milk-based snacks or desserts can attract more customers and increase your profit margin. Here are some steps to implement this strategy:

1. **Market Research**: Start by understanding what kind of milk-based snacks or desserts are popular in your area. You could do a quick survey with your current customers or observe what's selling well in nearby food businesses.
2. **Simple Recipes**: Choose simple recipes that require minimal additional ingredients and equipment. Puddings, custards, or simple ice creams could be good starting points.
3. **Pricing**: Price these items competitively while ensuring you still make a profit after considering the cost of ingredients, preparation time, and utilities.
4. **Promotion**: Use signage at your shop to promote these new products and consider offering samples to regular customers to get their feedback and generate interest.
5. **Quality Control**: Ensure the quality of your desserts is consistently high so that customers keep coming back for more.

Enter a number to learn more about one of these points on *milk-based snacks/desserts* or ask a new question.

Q10. 1

- A10. Absolutely, conducting market research is a crucial first step. Here are some actionable steps:
1. **Surveys**: Create simple surveys asking your customers about their favorite milk-based snacks or desserts and if they'd be interested in buying them from your shop. This could be done informally when they visit your shop or through a more structured approach like questionnaires.
 2. **Observation**: Visit other food businesses in your area that sell milk-based snacks or desserts. Note which items seem popular, how they're priced, and any unique selling points.
 3. **Community Engagement**: Engage with local community groups or online forums to understand what types of desserts people in Nairobi enjoy and are currently trending.
 4. **Trial Products**: Once you've gathered this information, select one or two products that seem promising and start offering them on a trial basis to gauge customer response.
 5. **Feedback Collection**: Collect feedback from the customers who try these new products - did they enjoy it? Would they buy it again? How much would they be willing to pay for it?

Remember: The goal of market research is not just to identify what products you should offer but also to understand how you can differentiate yourself from competitors.

Enter a number to learn more about one of these points on *market research* or ask a new question.

Example 2: Fast Food Joint Owner (Above Median)

In this conversation, the owner of a fast food joint located near a matatu terminus sought advice on differentiating their business in a competitive environment. The key strategies discussed included differentiating the menu with unique or special dishes, improving customer service, implementing a loyalty program, engaging in community activities, and maintaining high standards of cleanliness and hygiene.

The discussion delved into the specifics of establishing a loyalty program, with the AI assistant offering suggestions like a point-based system, special offers for regular customers, and birthday or anniversary rewards. The entrepreneur raised concerns about the cost implications of differentiating the menu and the complexities of community involvement. The AI assistant then suggested using affordable local ingredients, offering smaller portions, running specials during off-peak hours, using simple loyalty systems like punch cards, and engaging in low-cost community activities.

The entrepreneur was also concerned about the risks and customer reception related to changing the menu and implementing new strategies. The AI assistant recommended testing new dishes gradually and seeking customer feedback, especially from regulars. The importance of using both in-person and digital channels to collect reliable feedback was emphasized, considering that some customers are transient. The entrepreneur and the AI assistant also discussed the risks associated with implementing new business strategies, with the AI assistant suggesting thorough market research, pilot testing, financial analysis, risk assessment, and regular feedback review.

Q1. I have fast food joint that I run just adjacent to a matatu terminus. There are several such food joints at that particular place and it's fair to say competition is stiff. How can I overcome the competitive business environment and be ahead of the rest of the pack?

- A1. Thank you for sharing about your business. Operating a fast food joint in a highly competitive environment like a matatu terminus can indeed be challenging, but there are several strategies you could consider to stand out:
1. **Differentiate Your Menu**: Offer unique or special dishes that aren't available at the other joints. This could be traditional Kenyan foods with a twist, or it could involve using locally sourced ingredients that others don't use. The benefit is that customers will associate your place with these special offerings and choose you over competitors.
 2. **Improve Customer Service**: Prioritize excellent customer service to make people feel welcome and appreciated when they visit your joint. This might involve training staff on how to interact with customers positively, quickly serving orders, and handling any complaints promptly and professionally.
 3. **Loyalty Program**: Implement a loyalty program where customers earn points for every purchase which they can later redeem for free meals or discounts. It encourages repeat business as customers will want to accumulate more points.
 4. **Community Involvement**: Engage in local community events or sponsor local initiatives such as sports teams or school activities. This raises your profile within the community and makes people more likely to support your business.
 5. **Cleanliness & Hygiene**: Ensure high standards of cleanliness both inside the kitchen and in the dining area because hygiene matters greatly in food businesses.

Enter a number to learn more about one of these points on *improving competitiveness* or ask a new question.

Q2. 3

- A2. Great, let's delve into how you can implement a **Loyalty Program** in your fast food joint:
1. **Point-Based System**: Implement a simple point-based system where customers earn points for every shilling spent. Once they accumulate a certain number of points, they can redeem them for free meals or discounts. You could use simple tools like punch cards or digital apps that are inexpensive to manage.
 2. **Special Offers for Regular Customers**: Identify your regular customers and offer them exclusive deals or discounts. This not only rewards their loyalty but also encourages them to keep coming back.
 3. **Birthday/Anniversary Rewards**: Offer special treats or discounts to customers on their birthdays or anniversaries if they share this information with you. This personalized touch can make customers feel valued and appreciated.
 4. **Referral Program**: Encourage your loyal customers to bring in new clients by giving them incentives such as a discount or free meal when someone they refer makes their first purchase.
 5. **Feedback Reward**: Encourage customer feedback by offering small rewards for completing surveys about the quality of food and service at your joint.

Remember, the goal of any loyalty program is to make your customers feel valued so that they continue patronizing your business over others.

Enter a number to learn more about one of these points on *loyalty programs* or ask a new question.

- Q3. 1. **Differentiating my menu* means that I'm incurring a higher cost of preparing meals. This in turn pushes the menu prices upwards which may scare away the potential customers considering the fact that the potential customers are low or middle income earners. 2. About *loyalty programs*, I'm not well conversant with that field. 3. *Community involvement* may require adequate resources and may be capital intensive.*

A3. Thank you for your detailed response. Let's delve into these topics a bit more:

1. ****Differentiating Your Menu****: You're correct that introducing unique dishes may increase costs. However, this doesn't always have to result in higher menu prices. Here are some tips: [1] Focus on dishes that use affordable local ingredients but are prepared in a distinctive way. [2] Consider offering smaller portions at lower prices for those who want to try something new without paying a lot. [3] Run specials during off-peak hours to attract customers and offset the cost of unique dishes.

2. ****Loyalty Program****: A loyalty program is essentially a reward system for frequent customers. Here's how you could implement one with minimal costs: [1] Use simple paper punch cards where each purchase gets a punch, and after a certain number of punches, the customer gets a free meal or discount. [2] Keep track of repeat customers manually and offer them discounts or free items after they've made several purchases.

3. ****Community Involvement****: While it can be resource-intensive, there are ways to engage with the community without spending much money: [1] Offer your venue as a meeting place for local groups during slow business hours. [2] Sponsor small-scale local events like neighborhood cleanups by providing refreshments - this helps build goodwill within the community.

Enter a number to learn more about one of these points on *cost-effective strategies* or ask a new question.

- Q4. *If we focus on having a menu that stands out from what others are offering, by what margin may I lose, retain or increase my customer base. And what is the likelihood of having rewards for customer's referrals?*

A4. The impact of differentiating your menu on your customer base can be significant, but it's hard to quantify without specific data. However, here are some general insights:

1. ****Potential Increase in Customers****: Offering unique dishes can attract food lovers who are always looking for something new and exciting to try. If your new offerings resonate with the taste preferences of the local population, you could see a substantial increase in customers.

2. ****Retention of Existing Customers****: If you continue offering popular items while introducing new ones, you're likely to retain most of your existing customers. They might even appreciate the added variety.

3. ****Possible Losses****: There is always a risk that some customers may not like changes to the menu and choose other joints over yours. It's essential to introduce changes gradually and get feedback from customers before making any drastic alterations.

As for referral rewards:

4. Referral programs can be highly effective because people trust recommendations from friends or family more than advertising messages from businesses. If executed well, this strategy could lead to a significant increase in new customers without requiring large investments in marketing.

To implement these strategies:

A) For *Menu Differentiation*, start by researching what foods or flavors are trending locally or globally that align with your capabilities and resources. Test these out as specials before adding them permanently to your menu based on customer response.

B) For *Referral Rewards*, consider giving both the referring customer and the referred friend a discount or free item on their next purchase. Make sure it's easy for customers to share referrals by providing them with simple instructions when they make purchases.

Enter a number to learn more about one of these points on *menu differentiation* or *referral rewards* or ask a new question.

Q5. I'll need to contemplate on this two strategies and evaluate the outcomes both on a short and long term basis.

A5. Absolutely, taking the time to contemplate and evaluate potential outcomes is a crucial step in making strategic decisions for your business. Remember that both *menu differentiation* and *referral rewards* are strategies aimed at improving customer retention and attracting new customers, which are key aspects of business growth.

When evaluating these strategies:

1. Consider your resources: Do you have the capacity to introduce new menu items or manage a referral program effectively?
2. Listen to your customers: What do they want or need? Would they appreciate a more diverse menu or would incentives for referrals motivate them?
3. Monitor results: Once implemented, track the success of these strategies through increased sales, more repeat customers, or an increase in new customers.
4. Be ready to adjust: If something isn't working as well as expected, don't be afraid to make changes.

Remember that successful businesses are those that adapt and evolve based on their environment and customer needs.

Feel free to return if you have any other questions or need further clarification on any point related to improving your fast food joint's competitive advantage

Q6. Customer feedback when venturing into new ideas is the key to monitor the reception of a new menu and also the trust built through referral programs may work well with the business and the projected profit margins.

A6. Absolutely, you're on the right track. Customer feedback is invaluable when introducing new ideas like a different menu or a referral program. Here's how you can effectively use customer feedback and trust building:

1. **Feedback Channels**: Establish easy ways for customers to give feedback such as suggestion boxes at your joint, conducting informal surveys, or simply talking to them about their experience.

2. **Respond to Feedback**: Show customers that you value their opinions by acting on their suggestions where feasible and communicating changes back to them.
3. **Trust Building in Referral Programs**: Make sure the process is transparent and rewards are delivered as promised. This will build trust not only with the referring customer but also with the new ones coming in through referrals.
4. **Evaluate Profit Margins**: Regularly review your costs versus revenues after implementing these strategies to ensure they're contributing positively towards your profit margins.
5. **Pilot Test**: Before rolling out any major changes, consider doing a small-scale pilot test to gauge reception and adjust based on feedback received.

Remember, open communication with your customers helps foster loyalty which can lead to increased profitability in the long run.

Enter a number to learn more about one of these points on *customer feedback* and *trust building* or ask a new question.

Q7. *Together with my team, we shall deliberate on few important tips I've come across here and decide the way forward in best way to implement them.*

- A7. That sounds like a great plan. Collaborative decision-making often leads to more effective strategies and smoother implementation. Here are some tips for your deliberation process:
1. **Set Clear Goals**: Clearly define what you want to achieve with each strategy, such as increasing customer retention, attracting new customers, or boosting profit margins.
 2. **Assign Roles**: Assign team members specific roles in the implementation of each strategy based on their skills and responsibilities.
 3. **Plan for Feedback**: Plan how you will collect and analyze customer feedback once the changes are implemented.
 4. **Prepare for Adjustments**: Be prepared to make adjustments based on the outcomes of your initial efforts.
 5. **Schedule Follow-ups**: Schedule regular follow-up meetings to assess progress and make necessary adjustments.

Remember that it's important to have everyone on board with these changes so they can be effectively implemented and managed over time.

Enter a number to learn more about one of these points on *team deliberation* or ask a new question.

Q8. *Good insight. I'll come back here sooner with some more insights on the progress made if any.*

- A8. That's great to hear! Remember, the process of implementing new strategies often involves learning and adjusting along the way. Don't hesitate to return if you have any questions or need further advice as you progress. Wishing you and your team all the best in enhancing your business performance and profitability.

Q9. *Hi... If I'm running a food joint near a matatu terminus, selling fast food, snacks and beverages, how can I diversify the services I offer*

A9. Hello! Diversifying your services can be a great way to attract more customers and increase revenue. Here are some ideas:

1. **Delivery Service**: You could start offering delivery service for customers who may not have time to visit your food joint in person but want a quick meal.
2. **Catering Services**: Consider providing catering services for local events or businesses, which can significantly boost sales.
3. **Meal Deals**: Introduce special meal deals during specific hours of the day, such as breakfast combos or late-night snacks.
4. **Seasonal Menu Items**: Offering seasonal menu items that align with holidays or local events can attract new customers and keep regulars interested.
5. **Loyalty Program**: Implement a loyalty program where regular customers earn points they can redeem for meals or discounts.

Each strategy has its own benefits and methods of implementation, so consider what would work best for your specific business situation and resources.

Enter a number to learn more about one of these points on *service diversification* or ask a new question.

Q10. *How do I increase the returns on my cafeteria?*

A10. Understanding your business needs is crucial. I'm assuming your cafeteria is in Kenya, but could you provide more details? What's the size of your cafeteria, and where is it located? What's the competition like? Any specific challenges you are facing?

Based on general information, here are some strategies to increase returns:

1. **Improve Menu**: Review your menu to ensure it appeals to local tastes and preferences. Make sure you're offering dishes that not only taste good but also have high-profit margins. Experiment with different dishes and keep track of which ones sell best.
2. **Cost Control**: Manage food waste by implementing inventory management systems and portion control measures. This can help reduce costs and increase profitability.
3. **Customer Experience**: Enhance the customer experience by training staff in customer service skills, improving café ambiance or cleanliness, or even offering free Wi-Fi if feasible.
4. **Promotions & Loyalty Programs**: Implement a loyalty program or offer special promotions during slow periods to attract more customers.
5. **Community Engagement**: Get involved with the local community through sponsorships or events which can enhance visibility and build a loyal customer base.

Enter a number to learn more about one of these points on *increasing cafeteria returns* or ask a new question.

Q11. *I have tried to offer a new special meal menu in my restaurant and I must admit that the returns have been quite fair. Is it okay if I do a whole menu overhaul and start offering some different special meals altogether, even though it would make the price of each plate to be higher than before. I'm stuck between giving new menu or staying with current ones for the fear of losing the current clientele.*

A11. That's a great question and it's normal to feel this way when considering major changes in your business. Here are some strategies you can consider:

1. **Test the Waters**: Before overhauling the entire menu, consider introducing new special meals one at a time. This will allow you to gauge customer response without risking too much.
2. **Customer Feedback**: Ask your customers directly for their opinion on potential new dishes or even higher prices. You could use surveys or informal conversations.
3. **Market Research**: Look at what similar restaurants in your area are doing successfully, especially those with a similar clientele.
4. **Pricing Strategy**: Consider if there's room to slightly increase prices of current dishes instead of introducing more expensive ones that might deter customers.
5. **Special Events**: Host special events where only the new menu is available, like themed nights or tasting events, to introduce customers to these dishes and gather feedback.

Enter a number to learn more about one of these points on *menu overhaul* or ask a new question.

Q12. 1

A12. Great, you're interested in the strategy of testing the waters with new special meals. Here are some additional tips on how to do that:

1. **Pilot Dish**: Start by introducing one or two new dishes as a special offer and observe customer reactions.
2. **Promote**: Make sure your customers know about the new dish through table tents, menu boards, or even verbally when they're ordering.
3. **Feedback**: Ask for feedback directly from those who try the new dishes. This could be done informally by waitstaff or through a small comment card.
4. **Evaluate**: Track sales of the new dishes and compare them to your regular items. If they perform well, consider adding them permanently.
5. **Adjust**: Based on feedback and sales data, adjust the dish if needed before fully integrating it into your menu.

Enter a number to learn more about one of these points on *testing new special meals* or ask a new question.

Q13. *As for the customer feedback, sometimes it may not be reliable since some of the clients we host are passengers on transit so some opinion they may give may not apply to everyday's consideration.*

A13. You're correct, transient customers may provide feedback that doesn't necessarily reflect the preferences of your regular clientele. Here are some strategies to ensure you're getting reliable feedback:

1. **Target Regular Customers**: Make a point to solicit feedback from your regulars who understand your current menu and can provide insights into potential changes.
2. **Digital Feedback**: If you have an online presence (e.g., social media), use these platforms to gather feedback as people might be more open and detailed in their responses.
3. **Feedback Cards**: Leave these at tables or the checkout counter for customers to fill out before they leave. This will give you a wider range of opinions.

4. **Observation**: Watch what dishes get ordered most often and which ones come back with leftovers – this can be very telling about what's working and what isn't.

5. **Trial Periods**: As mentioned earlier, introduce new items slowly and see how they perform over a set period before making permanent decisions.

Enter a number to learn more about one of these points on *collecting reliable customer feedback* or ask a new question.

Q14. *A shift in business strategies may sometimes backfire. Implementing new ideas may fail the business often. How can I avoid the setbacks off trying new ideas and taking risks?*

A14. Trying new ideas and taking risks is a part of business growth, but it's important to manage these risks wisely. Here are some strategies:

1. **Market Research**: Before implementing a new idea, conduct thorough market research to understand its potential impact on your customers and competitors.

2. **Pilot Testing**: Instead of implementing changes across the entire business, try them in a small controlled environment first to gauge their effectiveness.

3. **Financial Analysis**: Consider the financial implications of the new strategy. Calculate potential returns against costs and assess if it's financially viable.

4. **Risk Assessment**: Identify potential pitfalls or challenges with your new strategy and develop contingency plans to address them.

5. **Feedback & Review**: Regularly review the performance of the new strategy, seek feedback from customers and staff, and be ready to make adjustments as needed.

Enter a number to learn more about one of these points on *risk management when trying new business strategies* or ask a new question.

E Survey Questions and Measurement

E.1 Measuring Business Performance

We measure business performance through standard survey questions that have been widely used in emerging markets (De Mel, McKenzie, and Woodruff, 2009). Specifically, we elicit measures of weekly profits, monthly profits, weekly revenue, and monthly revenue in each of the three pre-treatment and four post-treatment surveys using the following questions:

1. **Weekly Sales:** *Now we would like to ask you about the total sales across all of your businesses in the last 7 days. Please think about all the money that came into your businesses during the past 7 days, or all the money that you collected from customers BEFORE paying for any expenses and wages.*
2. **Monthly Sales:** *Now we would like to ask you about the total sales across all of your businesses in the last 30 days. Please think about all the money that came into your businesses during the past 30 days, or all the money that you collected from customers BEFORE paying for any expenses and wages.*
3. **Weekly Profits:** *Now we would like to ask you about the total profits across all of your businesses in the last 7 days. What was the total income the business earned during the last 7 days after paying all expenses including the wages of employees, but not including any income you paid yourself.*
4. **Monthly Profits:** *Now we would like to ask you about the total profits across all of your businesses in the last 30 days. What was the total income the business earned during the last 30 days after paying all expenses including the wages of employees, but not including any income you paid yourself.*

E.2 Explanation of Preferred Functional Forms

In our preanalysis plan, we defined four different variations of our performance index: 95% and 99% winsorization and a log-like inverse hyperbolic sine transformation $y' = \log(y + (y^2 + 1)^{1/2})$ of these two winsorizations. In the paper body, we focus on the raw performance index for two reasons.

First, recent work has shown that the inverse hyperbolic sine transformation (along with other log-like transformations) is highly sensitive to statistical artifacts like the units used to measure outcomes (e.g., measuring revenue in individual Kenyan Shillings versus measuring the same revenue in thousands of Shillings) and the presence of zeros (Chen and Roth, 2024; McKenzie, 2023). We observe this zero-sensitivity problem in our data. Figure A9 shows the distribution of our performance indices, winsorized at 95% (left) and 99% (right). The top row shows raw units and the bottom shows log-like units. For the log-like index, zeros create an extreme left tail with outcome values close to or above -10 standard deviations.

Second, for the purpose of studying performance gaps, it makes more sense to focus on raw changes as proportional changes can reflect radically different performance impacts depending on whether the entrepreneur comes from the top or bottom of the distribution.

E.3 Survey-based Mechanisms Questions and Analysis

This section provides an overview of a series of prespecified survey measures designed to explore potential mechanisms. We start by providing the verbatim survey questions for the seven sets of mechanisms:

Management practices (from McKenzie and Woodruff (2017); McKenzie, Osman, and Rahman (2023)): In the following set of questions, we will ask you about certain business practices that you may or may not be implementing in your business. Please only respond ‘Yes’ if you have implemented this business practice at least one time in the last 3 months. It is okay if you are not doing every activity we describe and not all the practices we ask about may make sense for all businesses to do. We are just trying to assess where your business stands now. Over the last three months...

[M1] ...have you visited at least one of your competitor’s businesses to see what prices they are charging? {No, Yes}

[M2] ...have you visited at least one of your competitor’s businesses to see what products your competitors have available for sale? {No, Yes}

[M3] ...you asked existing customers whether there are any other products they would like your business to sell or produce? {No, Yes}

[M4] ...have you talked with at least one former customer to find out why they have stopped

buying from your business? {No,Yes}

[M5] ...have you asked a supplier about which products are selling well in your business's industry? {No,Yes}

[M6] ...have you ever attracted new customers with a special offer? {No,Yes}

[M7] ...have you marketed or advertised in any form? {No,Yes}

[M8] ...have you made use of digital social media for marketing purposes? (e.g., Facebook, WhatsApp, Twitter, LinkedIn) {No,Yes}

[M9] ...have you attempted to negotiate with a supplier/vendor for a lower price on raw material/stock/inputs to your business? {No,Yes}

[M10] ...did you keep track of every sale made by the business (all money into the business)? {No,Yes}

[M11] ...did you keep track of every purchase/expenditure (all money out) made by the business? {No,Yes}

[M12] ...did you use records regularly to know whether sales of a particular product are increasing or decreasing from one month to another? {No,Yes}

[M13] ...did you know which goods, products, or services make the business the most profit per item selling? {No,Yes}

[M14] ...have you worked out the cost to your business for each of your main products or services you sell? {No,Yes}

[M15] ...did you have a written budget, which states how much is owed or should be spent each month on rent, electricity, equipment maintenance, transport, advertising, and other indirect costs to business? {No,Yes}

[M16] ...have you set targets for how much you will sell? {No,Yes}

Time management

[TM1] In the last seven days, how effectively did you feel you managed your time in running your business? {I managed my time running my business very poorly, Poorly, Neither poorly or well, Well, I managed my time running my business very well}

[TM2] In the last seven days, how many of your planned tasks were you able to accomplish? {I accomplished none of my planned tasks, A few of them, About half of them, Most of them, I accomplished all of my planned tasks}

[TM3] In the last seven days, how often did you find yourself working outside of your

planned hours or feeling rushed? {Never worked outside planned hours or felt rushed, Rarely, Sometimes, Often, Always worked outside planned hours or felt rushed}

Innovation

- [In1] In the past 30 days, have you considered any new products, services, or business process ideas for your business? {No (I have NOT considered any new products, services, or business process ideas), Yes (I HAVE considered new products, services, or business process ideas)}
- [In2] In the past 30 days, have you tested or experimented with any new products, services, or business process ideas for your business? {No (I have NOT tested or experimented with new products, services, or business process ideas), Yes (I HAVE tested or experimented with new products, services, or business process ideas)}
- [In3] In the past 30 days, have you implemented any new products, services, or business process ideas for your business? {No (I have NOT implemented with new products, services, or business process ideas), Yes (I HAVE implemented with new products, services, or business process ideas)}

Technology use

- [TE1] How frequently do you use technology (like computers, applications, software) in running your business? I never use technology (like computers, applications, software) in running my business {Rarely, Sometimes, Often, I very often use technology (like computers, applications, software) in running my business}
- [TE2] To what extent do you think technology has impacted your business performance like your growth in profits or customer base? {Technology has had no effect, Technology has had a small effect, Technology has had a moderate effect, Technology has had a large effect, Technology has had a very large effect}
- [TE3] In the last 30 days, have you started using any new technologies in your business? {No (no new technologies), Yes, one new technology, Yes, two new technologies, Yes, three or more new technologies}

Information: Quantity and quality

- [IQ1] In the last 30 days, how easy has it been for you to get advice, mentorship, or information for your business when you need it? {Very difficult to get advice, men-

torship, or information, Difficult, Neutral, Easy, Very easy to get advice, mentorship, or information}

- [IQ2] In the last 30 days, how would you rate the quality of advice, mentorship or information you received for your business (across all sources you receive advice, mentorship, and information from)? {The quality of advice, mentorship, and information I receive is Very poor, Poor, Neutral, Good, The quality of advice, mentorship, and information I receive is very good}

Information: Stigma and negative judgment

- [IS1] In the past 30 days, how often have you felt judged or looked down upon when asking for advice, mentorship, or information for your business? {I have always felt judged or looked down upon when asking for advice, mentorship, or information for my business, Almost always, Sometimes, Almost never, I have never felt judged or looked down upon when asking for advice, mentorship, or information for my business}

- [IS2] In the past 30 days, has fear of negative judgment ever prevented you from seeking advice or mentorship or information for your business? {Fear of negative judgment has always prevented me from seeking advice or mentorship or information for my business, Almost always, Sometimes, Almost never, Fear of negative judgment has never prevented me from seeking advice or mentorship or information for my business}

Information: Relevance of mentorship/advice

- [IR1] In the past 30 days, is the advice, mentorship, or information you have received directly relevant to your current business needs? {The advice I received is never relevant to my current business needs, Rarely, Sometimes, Often, The advice I received is always relevant to my current business needs}

- [IR2] In the past 30 days, how frequently have you found the advice, mentorship or information you received to be modern and up-to-date with current business trends and practices? {The advice I received is never modern and up to date, Rarely, Sometimes, Often, The advice I received is always modern and up to date}

Analysis of Survey-Based Mechanism Questions

This section provides an overview of the analysis of the families of survey outcomes outlined in the previous section:

1. Management practices (sixteen items)
2. Time management (three items)
3. Technology use (three items)
4. Innovation activities (three items)
5. Information: Relevance (two items)
6. Information: Access stigma (two items)
7. Information: Accessibility and quality (two items)

For each set of outcomes we create a simple aggregate index. Specifically, for each survey question within each outcome family, we first standardized the question relative to the control group. We then constructed the index by taking the simple average across these variables in each survey wave. The analysis of these mechanism questions is presented in [Figure A8](#). In this figure, we apply the same strategy used to identify average and heterogeneous treatment effects on business performance in the main text. Across these prespecified survey mechanisms, we observe no consistent patterns in terms of either average or heterogeneous treatment effects.

One possible explanation for the lack of treatment effects is a unique feature of the AI assistant: unlike standardized business training programs that provide identical content to all participants, the assistant's responses are tailored to the topics raised by each entrepreneur. This customization may lead to variation in mechanisms and outcomes across users that is not captured by our prespecified survey measures. The lack of clear patterns in the survey data motivates our shift to analyzing the text of entrepreneur-AI conversations to better understand the sources of heterogeneity in the main results.

F Ethical Considerations

In our field experiment, we found no overall impact on business performance from the treatment intervention. However, we did detect heterogeneous treatment effects with respect to initial performance levels; high performers experienced a performance increase due to AI advice, while low performers experienced a performance decrease. These unexpected negative results gave us pause. Given that our treatment appears to have led to worse performance for some entrepreneurs in our sample, in this section we outline why our experiment is ethical, above and beyond having received approval from and followed the policies of the UC Berkeley’s Institutional Review Board.

First, according to [Asiedu et al. \(2021\)](#), a key ethical consideration is policy equipoise, which involves meaningful uncertainty about the efficacy of each treatment arm. Indeed, if there was consensus that one treatment arm would be superior to the other, then it would *not* be ethical to randomize it unless the treatment was costly and therefore scarce. In the context of our experiment, there was uncertainty regarding participants’ benefits from each arm of the study. To the best of our knowledge, we are the first to test whether AI advice can increase entrepreneurial performance, especially in a context like Kenya. That said, recent research on generative AI and research on advising and mentoring between entrepreneurs suggested that an AI assistant had the potential to increase business performance. While we believed there were benefits, there was no consensus at the time we deployed the experiment that one treatment would be superior to the other. Given the presence of equipoise, we argue it was more ethical to test the impact of AI with a few hundred entrepreneurs in a well-designed study, rather than not testing and making policy decisions that could a) deny millions of entrepreneurs access to a useful tool if the treatment effect was positive or b) encourage adoption of a technology that hurts productivity by millions of entrepreneurs if the treatment effect was negative ([Glennerster and Takavarasha, 2014](#)).

Beyond the presence of equipoise, we also ensured all participants in our study were well-informed and free to decide whether to participate or not. Our participants were also free to decide whether to continue using our AI assistant, and it was easy to stop using it, consistent with the ethical standards around participant choice.

While our experiment revealed that one particular subgroup did not benefit from access to the AI assistant, negative treatment effects for subpopulations commonly occur in experiments conducted in the social and medical sciences (Karlan, Knight, and Udry, 2015). Given the growing use of generative AI, understanding the impact of such tools on society is especially valuable.

G Experimental Spillovers

In this appendix, we examine the plausibility of two types of spillovers that could impact our results: treatment access spillovers, where non-treated individuals might gain access to the AI assistant, and performance spillovers (Holtz et al., 2023), in which performance improvements (or declines) for one individual come at the direct cost (or benefit) of another.

Treatment Access Spillovers

The AI assistant’s WhatsApp contact number is publicly available, allowing anyone to text in. Without strict controls, this raises the possibility that untreated participants could interact with the AI assistant. To address this, we developed our AI assistant in such a way that it only responds to messages from individuals on a prespecified access list. Only members of the research team were able to modify this access list.

Performance Spillovers

Another potential concern is performance spillovers within our experimental sample. Specifically, the heterogeneous treatment effects we observe might be driven by high-performing entrepreneurs capturing profits or revenue from low-performing entrepreneurs in the same sample. We consider this unlikely because our recruitment Meta advertisement targeted a broad audience and drew participants from across Kenya, a country with over 50 million people and high Facebook penetration (Wamuyu, 2020). This recruitment strategy resulted in a geographically diverse sample. Figure A1 shows the distribution of respondents by Kenyan county, and highlights the wide variation in location.

Despite the fact that it is unlikely that spillovers would be affecting our experimental estimates, it *is* possible that performance spillovers occur between entrepreneurs in our sample and their competitors that are *not* in our sample. Keeping this in mind, the low level of technology saturation and infeasibility of the diffusion of our AI assistant are key to interpreting our results. If there were a greater level of overall generative AI saturation in Kenya, our results may not hold. For instance, it is conceivable that with high levels of generative AI adoption, firm practices could uniformly improve and/or converge; this could lead to greater price competition and lower firm profits. Conversely, widespread adoption of generative AI could drive positive performance spillovers stemming from the diffusion of best-use practices among businesses. While our experiment is unable to shed light on these dynamics, we consider this a fruitful area for future research.

H Text Analysis

This section describes our text data and analysis in detail. Subsection H.1 provides an overview of our three sources of text data (entrepreneur questions, AI responses, and business changes descriptions) along with summary statistics on these datasets. Subsection H.2 describes how we use GPT-4 to develop taxonomies for classifying entrepreneur questions into 10 business topics and AI responses into 12 topics, showing that low and high performers ask about similar topics and receive similar types of advice. Subsection H.3 describes our word- and embedding element-level analyses testing whether the content of the AI’s advice causally impacts the business changes made by entrepreneurs, including a bag-of-words validation of our embedding approach. Subsection H.4 details how we constructed treatment effect embedding vectors to systematically identify representative examples of how treated versus control entrepreneurs make different business changes. Finally, Subsection H.5 describes our quantitative validation tests using keyword matching and word frequency analysis to test the qualitative patterns identified through the treatment effect embeddings.

H.1 Overview and Summary Statistics

Our text analyses use three sets of data. First, we use data on the quantity and content of messages each entrepreneur sent to the AI assistant. These messages include questions asked of the AI, numeric followups (e.g. the number “3” in [Figure 2](#)), and non-business and other messages sent to the AI. Panel A of [Table A5](#) presents summary statistics at the individual message level. Given that many questions are uninterpretable fragments (e.g. “3”) that can introduce noise into the text analysis process, we had two human coders read through the entire set of 4,810 messages, identifying messages that were business-focused and had sufficient context for our text analysis methods. We filtered out messages that were numeric (e.g. following up with “1”), fragments and follow-ups (e.g. “tell me more,” “Can you explain how I might do 3?”), and a small number of messages not focused on business topics (e.g., a business owner who asked about how to quit smoking). Given that entrepreneurs ask a large number of follow-ups, we were left with 1,392 business-focused messages suitable for text analysis and topic classification. For our text analyses and when estimating word embeddings, we combine all business messages sent by each entrepreneur to the AI into a single text blob per entrepreneur, with summary statistics for these combined text blobs shown in Panel A of [Table A11](#).

Second, we use data on the AI’s responses to entrepreneurs. The AI responded to every message it received from entrepreneurs. These responses were often structured to provide 3-5 pieces of advice, as shown in [Figure 2](#). Panel B of [Table A5](#) provides summary statistics on the number of pieces of advice generated and message length. As with the entrepreneur messages, we concatenate all AI responses sent to each entrepreneur into a single text blob per entrepreneur (Panel B of [Table A11](#)) and use this combined text blob to estimate word embeddings for AI advice.

Third, we use data on business changes each entrepreneur describes making. As part of our post-treatment surveys, we asked entrepreneurs to describe changes they made to their business during the final three survey waves. Entrepreneurs who did not make changes were not required to provide descriptions. [Table 2](#) shows that neither treatment nor performance status predicts whether an entrepreneur describes making changes. The specific survey

questions were:

- Please describe in detail any new products that you have implemented in the last 30 days. If you have not implemented any new products, you can leave this question blank.
- Please describe in detail any new services that you have implemented in the last 30 days. If you have not implemented any new services, you can leave this question blank.
- Please describe in detail any new business processes or strategies that you have implemented in the last 30 days. If you have not implemented any new business processes or strategies, you can leave this question blank.
- In the last 30 days, which change that you made to your business has had the largest impact on performance (for example, sales and profits)? Please describe it in as much detail as possible.

As with the messages sent and advice received, we analyze the business changes made by concatenating all described changes for each entrepreneur into a single text blob and use this combined text blob to estimate word embeddings for business changes. Summary statistics for these combined business changes text blobs are shown in Panel C of [Table A11](#).

H.2 Question and Answer Topic Analysis

H.2.1 Taxonomy Generation for Business Messages

To understand the topics in the messages sent to and answers generated by the AI, we follow an approach outlined in [Choi et al. \(2023\)](#) and use GPT-4 to identify topics within the business-related questions posed by entrepreneurs to the AI assistant. This process has four stages:

Stage 1: We had two human coders read through the entire set of 4,810 messages, identifying messages that were business-focused and had sufficient context for topic classification. We filtered out messages that were numeric (e.g. following up with “1”), fragments and follow-ups (e.g. “tell me more,” “Can you explain how I might do 3?”), and messages not focused on business topics (e.g., a business owner who asked about how to quit smoking). Given that entrepreneurs ask a large number of follow-ups, we were left with 1,392 business-focused

messages suitable for topic classification.

Stage 2: Using this sample of business messages, we instructed GPT-4 to generate topics for classifying the messages in the corpus using the prompt below.

Listed below are queries submitted by SME entrepreneurs in Kenya to an AI Business Mentor designed to help them in their daily business operations.

The AI Business Mentor is designed to provide guidance and support to SME entrepreneurs in Kenya by answering their queries and providing them with valuable insights and advice. For each query presented, provide [num_topics] high-level business topics in less than 4 words each.

For example, if the query is “How can I improve my sales?”, the corresponding 4 topics could be [“Sales Strategy”, “Customer Acquisition”, “Marketing Campaigns”]

Return the topics for each query in a JSON format with the queryID, query and the corresponding topics.

The query list can be found below:

query.list

We ran this process multiple times, beginning with 3 topics per query and incrementally increasing to 20 topics per query. We then combined all topics generated across these runs to form a comprehensive set of 207 distinct topics. For robustness, we randomly selected approximately 10% of the topic assignments to messages and manually reviewed them to verify the topics were sensible.

Stage 3: We then took the set of 207 topics and instructed GPT-4 to merge or eliminate redundant topics, narrowing them down to a more concise list. We manually refined this list to remove overly specific topics, settling on a set of 10 topics:

1. Financial Management
2. Marketing Strategies
3. Risk Management
4. Customer Acquisition & Retention
5. Business Planning & Growth
6. Technology Integration & Digital Marketing
7. Inventory & Operations Management
8. Market Research & Analysis
9. Livestock & Farm Management
10. Entrepreneurial Opportunities & Business Ideas

Stage 4: Using these ten high-level topics, we instructed GPT-4 to classify each business message into one of these categories. Messages could only be classified into a single topic. We used the following prompt to classify these messages:

Listed below are queries submitted by SME entrepreneurs in Kenya to an AI Business Mentor designed to help them in their daily business operations. The AI Business Mentor is designed to provide guidance and support to SME entrepreneurs in Kenya by answering their queries and providing them with valuable insights and advice.

Provide a classification of each query using ONE high-level business topic from the list of topics below. For example, if the query is “How can I improve my sales?”, the corresponding topic could be “Marketing Strategies” or “Customer Acquisition & Retention.”

Here is the list of topics to choose from:

1. Financial Management
2. Marketing Strategies
3. Risk Management
4. Customer Acquisition & Retention
5. Business Planning & Growth
6. Technology Integration & Digital Marketing
7. Inventory & Operations Management
8. Market Research & Analysis
9. Livestock & Farm Management
10. Entrepreneurial Opportunities & Business Ideas

Here is the query list that you need to categorize: [query_list]

Return the topics for each query in a JSON format with the queryID, query and the topic.

We then used the topics generated from this procedure to create [Figure A10](#) which shows the distribution of topics by performance status. The topic distributions are very similar for high- and low-performers.

H.2.2 Taxonomy Generation for AI Mentor Responses

Following the process outlined above, we also generated a set of topics for the messages generated by the AI assistant in response to the entrepreneurs’ business messages. The result of the first three steps is a list of 12 topics:

1. Financial Management
2. Marketing Strategies
3. Business Growth & Expansion
4. Operational Efficiency
5. Agriculture
6. Customer Engagement
7. Starting a Business
8. Product Differentiation
9. Market Research & Analysis
10. Risk Management
11. E-commerce Platforms
12. Human Resources

[Figure A11](#) shows the distribution of response topics for low and high performers. The distribution largely mirrors what we see in [Figure A10](#). We again find no differences in the topics across the two groups.

H.3 Do Entrepreneurs Follow AI Advice?: Bag of Words Analysis

As a robustness check for our embedding-based approach for measuring whether entrepreneurs follow AI advice, we conducted a complementary “bag of words” analysis. Table A6 shows that when one of the 12,762 non-stop words in the AI’s responses is present in an entrepreneur’s advice text, it increases the probability of that same word appearing in the entrepreneur’s business changes text by 2.84 percentage points ($p < 0.0001$, 95% CI = [2.345; 3.327])—roughly a 15-fold increase given the 0.18 percentage point base rate. However, this approach rests on the strong assumption that words impact only themselves (e.g., “price” in advice affects only “price” in business changes, not related terms like “discount”) (Grimmer and Stewart, 2013). This “bag of words” assumption is clearly violated when, for example, advice about “pricing” leads entrepreneurs to describe “discounts.” Because of this, we prefer our embedding-based approach, which addresses these concerns by capturing semantic relationships across the full 3,072-dimensional vector space, ruling out edge cases where entrepreneurs use semantically similar but different words.

H.4 Technical Details: Treatment Effect Embedding Methodology

As mentioned in the main text, we developed a novel empirical strategy to identify the types of text changes that best represent the average differences between groups. This strategy involves first calculating the average difference in text embeddings between treatment and control groups for low performers and then for high performers. By comparing these embedding vectors, we capture how the AI-generated advice systematically changed the types of changes low performing entrepreneurs report making to their business. Specifically, we calculate the average business change embedding vector for low performing treated entrepreneurs, calculate the average business change embedding vector for low performing control entrepreneurs, and then compute the element-wise difference between treated and control averages (the “treatment effect embedding”). We then do exactly the same procedure for high performers.

Because these 3,072-dimension vectors are not easily interpretable on their own, we then

take pairs of control and treated business changes text, calculate the difference in the embedding vector, and identify individual text pairs that are most similar (using cosine similarity distance) to the average difference between groups. In [Table A9](#) and [Table A10](#) we qualitatively illustrate how treated versus control entrepreneurs differ in the types of business changes they make. Both tables show example pairs of treated versus control text sampled from the top 1% of most cosine similar text pairs to the high or low performer treatment effect embedding.

H.5 Quantitative Validation: Technical Implementation Details

For our word matching analysis we construct three measures. First, to test whether the entrepreneur mentions working with and learning from the AI, we take the raw business changes text and match it with the following regular expression pattern:

```
AI | [Cc]hat | [Gg] [Pp] [Tt] | [Aa] [Ii]  
[Mm]entor | [Aa]rtificial [Ii]ntelligence
```

We classify an entrepreneur who matches this pattern as having described working and learning from the AI.

Second, to test if the entrepreneur references the PDF guide we distributed to the control group, we test if the raw business changes text matches the following regular expression pattern:

```
[Gg]uide | [Pp] [Dd] [Ff] | [Mm]anual
```

Third, to test if the entrepreneur offered discounts or invested in advertising we match whether the raw business changes text matches the following regular expression:

```
[Ff]ree | [Dd]iscount | [Rr]educ | [Oo]ffer | [Aa]dver
```

We deliberately do not match against “marketing” as many entrepreneurs describe how they tried non-paid marketing and in [Table A10](#) we see no differences in the use of this term across the groups; instead the differences appear to be about whether the entrepreneur invests in

paying for a billboard or online ads. We include “free” because many entrepreneurs mention offering free delivery and other services that implicitly serve as a discount.¹⁷

We estimate linear probability models regressing the binary indicator for whether an entrepreneur’s text matched each regular expression on treatment assignment and baseline performance status. For these measures, we assume that entrepreneurs who do not describe making any business changes do not match any of the regular expressions, since activities such as working with the AI, building on the guide, or implementing pricing strategies necessarily require making business changes. [Table A7](#) presents regression estimates for AI mentions (Panel A), ILO business guide references (Panel B), and discounting/advertising strategies (Panel C).

Our final measure counts the number of uncommon words used in an entrepreneur’s business changes text. To create this measure, we removed capitalization, tokenized words, and removed stop words from the business changes text. We then classified words as uncommon if they appear in 5 or fewer business change descriptions. These words include “150cc,” “arcade,” “biodegradable,” “detergent,” “headphones,” “peanut,” “rabbit” and “uganda.” Common words include “advertising,” “customers,” “market,” “price,” and “vegetables.”

[Table A8](#) presents Poisson regressions testing whether treatment and performance status predict the count of uncommon words used. We use Poisson regressions since the outcome is a highly skewed count variable. Since entrepreneurs who write longer responses will mechanically use more uncommon words, we control for the total word count in the business changes text. We restrict our analysis to entrepreneurs who provided at least one analyzable word in their business changes text.

¹⁷While these regular expressions provide a systematic approach to quantifying the patterns observed in [Table A9](#) and [Table A10](#), they necessarily involve some measurement error inherent to keyword-based classification methods.