

The Uneven Impact of Generative AI on Entrepreneurial Performance

Nicholas G. Otis*

Rowan Clarke

Berkeley Haas

Harvard Business School

Solène Delecourt

David Holtz

Rembrand Koning

Berkeley Haas

Berkeley Haas

Harvard Business School

[\[Click here for latest version\]](#)

There is a growing belief that scalable and low-cost AI assistance can improve firm decision-making and economic performance. However, running a business involves a myriad of open-ended problems, making it hard to generalize from recent studies showing that generative AI improves performance on well-defined writing tasks. In our field experiment with 640 Kenyan entrepreneurs, we assessed the impact of AI-generated advice on small business revenues and profits. Participants were randomly assigned to a control group that received a standard business guide or to a treatment group that received a GPT-4-powered AI business mentor via WhatsApp. We are unable to reject the null hypothesis that generative AI access has no impact, but are able to rule out the large effect sizes reported by other studies of generative AI's economic impact. Our overall null result masks treatment effect heterogeneity with respect to the baseline business performance of the entrepreneur: our point estimates suggest that high performers benefited by just over 15% from AI advice, whereas low performers did about 8% worse with AI assistance. Exploratory analysis of WhatsApp interaction logs shows that both groups sought the AI mentor's advice, but that low performers did worse because they sought help on more challenging business tasks. Our findings highlight the potential and limitations of generative AI to enable entrepreneurs across the globe.

*Otis: notis@berkeley.edu (Corresponding author); Clarke: rclarke@hbs.edu; Delecourt: solened@berkeley.edu; Holtz: dholtz@haas.berkeley.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 Channing Jang for their exceptional support in our field operations, and Sara Dahiya, Brian Mutisyo, 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 mentor, 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's Conference on Digital Experimentation (CODE), and the Conference on Field Experiments in Strategy (CFXS). 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 (The White House, 2022; Agrawal, Gans, and Goldfarb, 2023; Eloundou et al., 2023; Hui, Reshef, and Zhou, 2023). Much of this recent work is driven by the belief that conversations with large language models (LLMs) can yield helpful assistance, feedback, and advice, ultimately improving firm performance and growth (Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023; Noy and Zhang, 2023; Peng et al., 2023; Kumar et al., 2023). Given the substantial variation that exists in both worker and firm productivity, both within and across countries (Bloom and Van Reenen, 2007; Bartelsman, Haltiwanger, and Scarpetta, 2013), the emergence of nearly zero marginal cost generative AI “mentors” 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 (McKenzie and Woodruff, 2017; McKenzie, 2021; 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 more quickly, including press releases, ad copy, consulting memos, and customer support messages (Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023; Doshi and Hauser, 2023; Noy and Zhang, 2023; Chen and Chan, 2023). However, it remains unclear whether the benefits of such AI feedback generalize to the broader set of tasks that firms engage in. Beyond these textual 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). Even for early-stage or small business entrepreneurs, the sheer multitude of tasks involved in running a business greatly increases the complexity of effectively integrating AI assistance into business activities (Lazear, 2004; Anderson and McKenzie, 2022). AI business advice also differs from AI code or writing assistance in that the advice cannot be “copied and pasted” and instead must be implemented by the entrepreneur. As a result,

while the advice provided by an AI tool might be useful in theory, the entrepreneur may lack the complementary knowledge, skills, and resources to correctly act on that agent’s recommendations (Brynjolfsson and Hitt, 2000; Brynjolfsson, Rock, and Syverson, 2021; Dell’Acqua et al., 2023). The plethora of tasks that entrepreneurs face may also result in inattention and failure to seek out advice or acquire new information, potentially limiting the benefits of even the most advanced AI-generated advice (Hanna, Mullainathan, and Schwartzstein, 2014; Kim, 2023).

Even when seeking advice from peers or mentors, it is often not obvious to entrepreneurs which tasks they should ask for help with, how to formulate effective questions to get useful feedback, or how to interpret and act upon the advice received (Bryan, Tilcsik, and Zhu, 2017; Camuffo et al., 2020; Agrawal, Gans, and Stern, 2021; Dimitriadis and Koning, 2022). These issues have the potential to be exacerbated when seeking advice from AI systems. Furthermore, even if the entrepreneur’s question is well-crafted, it is unclear how useful advice from AI will be in practice. Given the myriad tasks entrepreneurs undertake, it is plausible that AI advice generates a mix of useful answers, ineffective recommendations, and potentially detrimental advice that could hinder firm performance (Dell’Acqua et al., 2023). If most tasks require personalized advice to be useful, then AI feedback is likely to be especially valuable given its ability to contextualize responses. However, if most tasks are standardized, then AI will help, but potentially no better than the common advice an entrepreneur might get from reading a book or browsing the web. It is also possible that most business tasks are tacit and not yet codified as text, in which case AI tools may lack the relevant training data to provide useful answers to an entrepreneur’s questions (Autor, 2014). Of greatest concern, generative AI might yield overconfident or flawed recommendations for many real-world business tasks (Moore, 2022; Dell’Acqua et al., 2023), leading entrepreneurs to implement “solutions” that worsen rather than improve firm performance.

In this paper, we measure the impact of generative AI on the business performance of emerging market entrepreneurs by conducting a randomized field experiment with 640 Kenyan entrepreneurs. In the experiment, treated participants were randomly assigned to receive access to a GPT-4-powered AI business mentor via WhatsApp, whereas control participants received a standard business guide. We are unable to reject the null hypothesis

that on average, access to generative AI did not have an effect on business performance relative to the control. Using an equivalence test (Simonsohn, 2015; Lakens, 2017), we can reject all effects that would have been detectable by Noy and Zhang (2023). Although our overall result is null, this result masks substantial heterogeneity with respect to pre-treatment firm performance; “low performers” did worse with access to the tool, whereas “high performers” did better. Analysis of our generative AI tool’s interaction logs suggests that these heterogeneous effects are driven by differences in the types of tasks for which the two sub-populations sought assistance; low performers sought help on more challenging business tasks.

Beyond identifying the impact of generative AI in our specific research setting, our study also makes multiple contributions to the conversation regarding the technology’s broader economic impacts. First, our research sheds light on whether and when AI advice can improve entrepreneurial learning and firm productivity, both of which are core concerns for both managers and policymakers. A large and diverse literature shows that business training programs can lead to modest but meaningful improvements in firm productivity (McKenzie, 2021). For the high performers in our sample, our point estimates suggest that access to generative AI may have driven improvements that were comparable in magnitude, suggesting that generative AI offers an opportunity to scale personalized business mentorship and training, but at a fraction of the cost of existing programs.¹ Second, understanding if 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 those studied in prior work (Sato et al., 2023).

2 Experimental Design

We measure the impact of AI advice on entrepreneurial performance by running a field experiment with 640 Kenyan entrepreneurs running small and medium businesses. We selected this research context because while recent experiments document that entrepreneurs

¹Similar to recent work highlighting the cost-effectiveness of text message interventions (Fabregas et al., 2024), our mentor costs a few dollars per participant whereas most training programs cost hundreds of dollars (USD) per participant (McKenzie, 2021).

in developing economies greatly benefit from the context-specific and wide-ranging nature of human feedback (Brooks, Donovan, and Johnson, 2018; Cai and Szeidl, 2018; Chatterji et al., 2019), this context has been almost entirely overlooked in recent research on the potential productivity impacts of generative AI (Björkegren, 2023). Because obtaining personalized advice is especially challenging in developing economies, the benefits of AI advice in these contexts might be especially large (Dimitriadis and Koning, 2022, 2023).

Specifically, we developed an AI “mentor” for Kenyan entrepreneurs using GPT-4, a large language model released by OpenAI in March 2023 (OpenAI, 2023). ChatGPT, a simple Q&A interface for interacting with LLMs released by OpenAI, rocketed to an estimated 100 million monthly active users within two months of its launch (Hu, 2023). Building on the popularity of ChatGPT’s chat interface, we developed our own AI tool (hereafter the AI mentor) that can be accessed through WhatsApp. Development of this AI mentor took place over four months and involved extensive user testing by the authors, research assistants, and small business owners in Kenya (see Appendix D for more details on the AI mentor development process). We selected WhatsApp because it is used by nearly 90% of people in Kenya (Wamuyu, 2020) and because of the low cost of sending text over WhatsApp relative to SMS texts. Figure 1 depicts an example participant interaction with the AI mentor.

Interacting with the AI mentor differs from interacting directly with an LLM such as GPT-4 in three ways. First, we developed a system prompt that provided the AI with additional context about the Kenyan small and medium businesses that would use our app, which increased the odds that the advice provided to entrepreneurs by the AI mentor was contextually relevant.² Second, we instructed our AI mentor to generate three to five practical pieces of advice in response to each question posed by the entrepreneur, with each piece of advice accompanied by an explanation of its benefits and implementation details. Finally, to encourage entrepreneurs to engage further with the advice they received, the system prompt instructed the AI mentor to structure responses in such a way that users could quickly and easily ask for more information about each piece of advice. The diverse

²For example, if asked about how to raise capital, GPT-4 may recommend raising venture capital funds; a financing strategy that is out-of-scope for the small and medium businesses in our study. Instead, our system prompt would lead the AI mentor to focus on contextually relevant alternatives, like approaching family and friends or building a “chama,” an informal cooperative society.

ways that entrepreneurs interacted with the AI mentor, as well as the level of personalization provided by the tool are further highlighted in Appendix H, which presents full chat logs for two entrepreneurs in our sample. In one conversation, a restaurant owner is considering changing the menu and asking for assistance thinking through the possibilities and sources of uncertainty involved in making this decision. In another conversation, a business owner selling wholesale and retail milk is asking how to expand their product offerings to increase profits. Other conversation topics across our full sample include how to motivate employees, the best way to deploy capital when expanding a store, tips for hatching and raising healthy chickens, and how to deal with bankruptcy.

Our AI mentor is one of many AI tools that have recently been developed to assist entrepreneurs and firms (Baxter and Schlesinger, 2023). However, despite the popularity of these tools, there is scant evidence of their causal impact. This could be in part because measuring the causal impact of AI assistance on business performance presents challenges. Conceptually, defining the appropriate counterfactual for the AI mentor is non-trivial because AI-based interventions can offer multifaceted benefits. Some of these benefits can also be obtained with non-AI interventions, whereas others cannot. In our context, the AI mentor provides personalized advice but also an opportunity for the entrepreneur to reflect on their business needs. As a result, the treatment effect of access to the AI mentor relative to an unassisted control might simply reflect the fact that giving entrepreneurs any form of assistance nudges them to think more about how to improve their business. Empirically, measuring business performance in developing economies almost always involves surveying participants, which raises concerns that giving entrepreneurs an AI mentor might lead to “demand effects,” i.e., changes in the behavior of participants due to cues about what is considered appropriate and/or desirable (Zizzo, 2010).

To address these conceptual and empirical issues, we tested our AI tool against a control group that received a placebo-like intervention: business guides designed specifically for entrepreneurs operating small and medium businesses in developing economies (International Labor Organization, 2015). By providing control participants with access to business training materials, we isolate the effect of access to an AI mentor, net of the impact of simply providing a resource aimed at helping entrepreneurs improve their businesses. Previous

research shows that providing business training texts has no impact, as asynchronous online or offline materials tend to be rarely used (Cusolito, Darova, and McKenzie, 2023; Mehmood, 2023; Davies et al., 2024).³ Equally as important, under the assumption that the demand effects induced by the business training guides are equivalent to those induced by the AI mentor, treatment effects should be attributable to actual changes in performance. This helps rule out the possibility that any changes we observe are merely due to demand effects, as both groups received a tool to help them improve their businesses.

We recruited entrepreneurs into our study over the Meta ad platform in partnership with the Busara Center, a Kenyan research organization, starting in May 2023.⁴ Our recruitment strategy involved running ads on Facebook to invite entrepreneurs to a short paid survey (Figure A1). All entrepreneurs who responded to our ads were required to pass basic attention checks and take part in three rounds of pre-treatment surveys delivered over the course of three months, which helped us reduce post-treatment attrition and ensure valid causal inference. Appendix E describes our surveys, which asked about firm profits, revenues, and management practices, among other measures. Our final sample includes 640 Kenyan entrepreneurs who completed all three pre-treatment surveys. The median entrepreneur in this sample was 26 years old, had been running their business for one year, and held a college degree (Table A1). Our sample reflects the heterogeneity present amongst Kenyan small and medium businesses, with entrepreneurs running businesses from fast-food joints to poultry farms to cybershops⁵ across Kenya (see Figure A2 and Figure A3). Pre-treatment performance in our sample ranged from monthly profits of 2,650 Kenyan Shillings (5th percentile; about \$15 USD) to over 55,000 Kenyan Shillings (95th percentile; about \$340 USD).

Following the final pre-treatment survey wave, the entrepreneurs in our sample were block-randomized into treatment and control, with entrepreneurs stratified based on their gender and pre-treatment business performance (Appendix C). Treated entrepreneurs received free and unlimited use of the WhatsApp-based AI mentor, along with regular re-

³While rarely read, these training guides can benefit entrepreneurs when they are pushed to read or learn the content (Dalton et al., 2021). This suggests that, compared to a pure control, our design will if anything yield more conservative estimates.

⁴See Figure 2 and Appendix C for more details on the experiment timeline and sample recruitment process, respectively.

⁵Cybershops provide a range of services ranging from computer access, internet browsing, printing, scanning, photocopying, and computer repair and technical support.

minders to use the tool. Control participants were provided easy access to the aforementioned business training guides and were also sent regular reminders to use the guides (Appendix D). Control participants were not provided any information about generative AI, nor the existence of the AI mentor. Entrepreneurs in the two groups were comparable in terms of both performance and the other characteristics we measured prior to treatment (Table A1). Of the 640 participants who were randomized into control or treatment, 634 (99%) completed at least one post-treatment survey and 622 (97%) completed all four post-treatment surveys deployed over the two months following treatment (Appendix C).

Firm performance exhibits considerable variability in emerging markets (Fafchamps et al., 2012; Anderson, Lazicky, and Zia, 2021). In light of this heterogeneity, we pre-registered several steps to improve the statistical precision and credibility of our experimentally-estimated causal effects (see Appendix E for more details on the pre-registration). First, our outcome variable is an index that combines standardized measures of weekly and monthly revenue and profits to reduce noise in our dependent variable. Second, we analyze our data using a variant of simple ordinary least squares (OLS) regression that conditions on the average pre-treatment performance data collected from participants in the three pre-treatment survey waves (McKenzie, 2012). Third, we control for additional pre-treatment variables using a double-LASSO approach (Belloni, Chernozhukov, and Hansen, 2014).⁶ Fourth, we pool data from all four of our post-treatment periods to further increase statistical power. We detail the construction of our outcome variables and our econometric strategy in Appendix E.

3 Results

Our main results are shown in Figure 3. Using the regression specification outlined in our pre-analysis plan, we estimate a null average treatment effect (ATE) of AI access on business performance. This result is robust to whether we winsorize performance at the 99% level ($\delta = 0.05$ standard deviations (s.d.), $p = 0.36$; 95% confidence interval (CI) = $[-0.06, 0.16]$)⁷

⁶The sets of covariates selected by the double-LASSO for each of our main analyses are reported in Table A4.

⁷All p -values reported in the paper are from two-tailed tests. In Appendix subsection E.5, we discuss and report adjustments for multiple comparisons. Unless otherwise noted, estimates in the main text are not adjusted.

or the 95% level ($\delta = -0.01$ s.d., $p = 0.92$, 95% CI = [-0.14,0.07]), and is also robust to the exclusion of 18 participants who did not complete our entire post-treatment survey panel (95% winsorization: $\delta = -0.01$ s.d., $p = 0.86$, 95% CI = [-0.14,0.07]; 99% winsorization: $\delta = 0.05$ s.d., $p=0.34$, 95% CI = [-0.06,0.16]) (see Figure A4). To interpret this null result, we conduct exploratory post-hoc equivalence tests with the aim of rejecting the presence of treatment effects greater than or equal to some minimum effect size threshold (Simonsohn, 2015; Lakens, 2017). Building on Noy and Zhang (2023),⁸ we first set this threshold to 0.26 s.d., which is the minimum detectable effect size that their study can reliably detect, and strongly reject the presence of a treatment effect greater than or equal to this threshold (threshold = 0.26, $Z = -3.89$, $p = 0.00005$).⁹ Although recent studies of AI’s impact on business performance have reported effect sizes approaching 0.5 s.d. (Noy and Zhang, 2023; Brynjolfsson, Li, and Raymond, 2023), these estimates are roughly five-times larger than the estimated impact of human-to-human business training (McKenzie, 2021). Thus, to further contextualize our null result, we next conduct a second exploratory equivalence test with the minimally important treatment effect threshold set to 0.09 s.d., which is the minimum detectable effect size reported in a recent study of personal initiative training by Campos et al. (2017). We are unable to reject the presence of treatment effects above this lower threshold (threshold = 0.09, $Z = -0.69$, $p = 0.28$). Taken in tandem, these equivalence test results indicate that while we can rule out the presence of the large treatment effects found in recent studies of generative AI, we cannot rule out smaller treatment effects on the order of those reported in studies of human-to-human business training in developing economies.

Although we fail to reject the null hypothesis of no overall effect, this null masks opposing heterogeneous treatment effects. Again following our pre-registration plan, we split our

⁸We focus on Noy and Zhang (2023) for two reasons. First, it is currently, to our knowledge, the only peer-reviewed and published experiment testing the impact of generative AI on business performance. Second, recent working papers including Brynjolfsson, Li, and Raymond (2023) and Dell’Acqua et al. (2023) do not provide sufficient information to convert their treatment effect estimates and confidence intervals into standard deviations, preventing us from making direct comparisons.

⁹An alternative way to gauge the precision of our estimate is to focus on the the ex post minimum detectable effect size (MDE) of our and prior studies directly. The MDE for a study with 80% power and a 5% rejection rate is simply the standard error multiplied by 2.8 (Imbens and Rubin, 2015). The standard error reported in Table A3 Model 2A is 0.054, which implies our study has an ex post MDE of 0.151. Recent experiments conducted in developing economies have roughly similar MDEs for performance outcomes, ranging from 0.087 s.d. (Campos et al., 2017) to 0.15 s.d (Dimitriadis and Koning, 2022).

sample of entrepreneurs based on pre-treatment performance as defined using the same performance index measure that serves as our primary outcome variable.¹⁰ For those with below-median pre-treatment performance (hereafter “low performers”), we find evidence that access to the AI mentor reduced average business performance ($\delta = -0.09$ s.d., $p = 0.007$, 95% CI = [-0.15,-0.03]; see Figure 3), but are unable to reject the null hypothesis of no impact on median business performance ($\delta = -0.03$ s.d., $p = 0.210$, 95% CI = [-0.07, 0.01]; see Figure 4 panel C).¹¹ Conversely, for entrepreneurs with above-median pre-treatment performance (hereafter “high performers”), we are unable to reject the null hypothesis of no impact on average business performance ($\delta = 0.19$ s.d., $p = 0.07$, 95% CI = [-0.02,0.38]; see Figure 3), but do find evidence that access to the AI mentor increased median business performance ($\delta = 0.10$ s.d., $p = 0.03$, 95% CI = [0.01, 0.18]; see Figure 4 panel D).¹²

Figure 4 provides insight into why we detect a mean, but not median effect for low performers, and vice-versa for high performers. In both cases, there are noticeable differences between the treated and control empirical cumulative distribution functions (eCDFs) (panels A and B). However, for low performers, the negative shift in the treatment eCDF is almost entirely above the median change in business performance. This leads to a statistically significant ATE, but a null quantile treatment effect (QTE) at the median. In the case of high performers, there is consistently a positive shift in the treatment eCDF; however, a substantial fraction of high performers in both treatment arms have very low (≤ -0.97) or very high (≥ 1.61) residualized performance indices in the post-treatment period. This leads to a noisy (and thus, null) ATE, but a statistically significant QTE at the median. These dynamics are made even more clear after inspecting the QTEs at each decile for both low performers (Figure 4 panel C) and high performers (Figure 4 panel D). In the case of low performers, all QTEs for deciles at or below 60% are null and relatively tightly estimated,

¹⁰Our pre-registered analysis plan included testing for differences in average business outcomes, but not median business outcomes. Figure 4 highlights why we include non-pre-registered analyses of quantile treatment effects alongside the pre-registered average outcomes. Following this pre-registration plan, we also tested for heterogeneous treatment effects with respect to both prior ChatGPT use (measured prior to the experiment) and gender, as female entrepreneurs in developing economies often face additional constraints compared to men, such as childcare (Delecourt and Fitzpatrick, 2021). We find no evidence of heterogeneous treatment effects on either dimension.

¹¹Appendix F discusses the ethical considerations for our study given that we detect this negative effect.

¹²See Table A3 for robustness checks of the results presented in Figure 3; Appendix G demonstrates that our results are not an artifact of spillovers between participants.

whereas QTEs for higher quantiles become increasingly large and statistically significant. In the case of high performers, all QTE point estimates are positive, but QTE estimates for higher quantiles are estimated with increasing levels of imprecision.

Overall, our results differ from those of prior studies not only because we rule out large positive effects of AI assistance such as those detected by Noy and Zhang (2023), but also because we report heterogeneous treatment effects with respect to pre-treatment performance that point in the opposite direction of those reported by prior work on the economic impacts of generative AI (Noy and Zhang, 2023; Peng et al., 2023; Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023). Moreover, the difference we observe between the estimated ATEs of AI for high and low performers is non-trivial and statistically significant ($\delta = 0.267$ s.d., $p = 0.013$, 95% CI = [0.06, 0.48]). To put our point estimates in perspective, our negative ATE point estimate among low performers is equivalent to an approximately 8% drop in profits or revenue among low performers, while our positive ATE point estimate among high performers is equivalent to an approximately 15% increase in profits or revenue among high performers.¹³ Given the apparent inconsistency between our results and those reported by recent studies of generative AI, we next present a series of exploratory analyses conducted with the aim of identifying mechanisms that explain the heterogeneous treatment effects we observe.

One major difference between our field experiment and prior research on the economic impacts of generative AI is the level of discretion participants were granted in how and when they used AI assistance. In previous experiments, experimenters narrowly constrained the tasks participants completed, the extent to which they engaged with the AI tool (Brynjolfsson, Li, and Raymond, 2023), and/or the length of time that they were granted use of generative AI (Noy and Zhang, 2023; Dell’Acqua et al., 2023). In contrast, the entrepreneurs in our experiment had much more discretion over how they used AI assistance. For instance, they were able to choose how many questions to ask the AI mentor, and the importance of the tasks on which they requested assistance. They could also ask for assistance on a wide range of topics, including, but not limited to financing, marketing, operations, and farming

¹³For details on these calculations see Table A11.

practice.¹⁴ The discretion afforded to participants in our study suggests that differences in AI usage between low and high performers could explain the heterogeneous treatment effects that we observe in our sample.

To understand whether this is the case, we first test for differences in the quantity and quality of business-related questions that treated entrepreneurs sent to the AI mentor.¹⁵ Figure 5 shows the distribution of number of business questions asked by both high and low performers during our experiment. Both distributions are quite skewed, with roughly 15% of entrepreneurs not asking a single business question and a small number of entrepreneurs asking one nearly every other day. Among both high and low performers, those that did ask at least one business question asked roughly one a week. Overall, the two distributions are extremely similar; although on average, high performers sent 0.59 more messages than low performers, we fail to reject the null hypothesis of no difference in the number of questions asked ($\delta = 0.59$, $p = 0.33$, 95% CI = [-0.60, 1.78]). We also test for differences in the human-evaluated quality (measured on a three-point scale) of questions asked by low and high performers, and the length (in words) of the questioned asked by the two groups. In both cases, we fail to reject the null hypothesis that there is no difference between the two groups ($\delta = -0.07$, $p = 0.294$, 95% CI = [-0.21; 0.06] and $\delta = -0.11$, $p = 0.136$, 95% CI = [-0.26, 0.04] respectively; see Table A13). In other words, we do not find evidence that the negative returns to AI assistance for low performers are explained by differences in the number, quality, or length of business questions that they asked the AI mentor relative to

¹⁴Consistent with this argument, we conduct a pre-registered analysis of the impact of the AI mentor on a number of pre-specified questions designed to identify potential treatment effect mechanisms. We find little evidence that the AI mentor led to consistent changes in our survey measures of management practices, time management, technology use, innovation, or information-seeking behavior (Figure A9). After accounting for the multiple comparisons made across these seven indices, the lowest p -value that we observe is 0.28. However, it is worth noting that if only 10% of our sample asked about management, 10% about time use, 10% about technology, and so on, then we would be severely underpowered to detect average treatment effects along any of these individual dimensions. More broadly, the personalized nature of AI suggests that social scientists may need to rethink how they identify the underlying mechanisms driving treatment effects when interventions are personalized and/or algorithmic in nature.

¹⁵While most questions sent to the AI mentor were business-focused, some entrepreneurs also asked the mentor about non-business topics (e.g. how to quit smoking) or sent in text by mistake (e.g. messages intended for someone else). The entrepreneurs also sent in questions of varying quality, sometimes writing full sentences and other times writing no more than a couple of cryptic words. Appendix I outlines the process by which two human coders evaluated the focus and quality of each question sent to the AI mentor. Our final dataset includes just over 1,300 business-related questions and excludes the numerous follow-up interactions entrepreneurs have with the mentor after asking an initial question.

high performers.

Although we do not detect differences in the quantity or quality of business questions asked by high and low performers, this does not mean that the questions asked by the two groups do not differ meaningfully; entrepreneurs in our study also had the freedom to choose the topics they discussed with the AI mentor. Given that, by definition, low performers had weaker revenues and profits than high performers prior to our treatment intervention, one possibility is that low performers simply sought advice on especially difficult tasks. These topics and problems may be well beyond what an AI mentor—or even a human mentor—can help with, in part because these problems often require financial capital and/or other complements to solve. For instance, a firm facing stiff competition or a farm experiencing a drought might struggle even after receiving extremely high-quality advice. Even changes that had a neutral impact on growth and/or sales could have resulted in lower profits and worse performance, given that many of the AI mentor’s recommendations were costly to implement. We use word embedding methods to develop a principled and scalable measure of the extent to which each question in our dataset is focused on an especially difficult business problem.¹⁶ Appendix J documents our approach in detail. In short, we map each question onto a standardized uni-dimensional *challenge-focused* measure. We do so by calculating the relative closeness of each question’s word embedding vector to the word embedding vectors for two different benchmark questions, one of which describes a business in dire straights and one of which describes a business facing few challenges.

Our *challenge-focused* measure has strong face validity. Questions that are rated as more challenge-focused tend to be about difficult problems that are unlikely to be fixed using an AI mentor’s recommendations.¹⁷ Some of the most challenge-focused questions include:

1. “currently facing loses in my show shop due to low demands rendering business bankrupt. please advise” (99th percentile)
2. “hi, i have a beauty shop behind [REDACTED] . at first it was the best selling shop around but now it’s nearly the last. what could be the problem and how do i solve it?” (95th

¹⁶Word embeddings map text into high dimensional vectors (Mikolov et al., 2013) and have been used to measure semantic differences ranging from whether a startup idea is more likely to benefit women to how cultural associations have evolved over the twentieth century (Cao, Koning, and Nanda, 2023; Kozlowski, Taddy, and Evans, 2019).

¹⁷Table A14 shows above-median-length example questions pulled from the 50th to 100th percentiles of the *challenge-focused* distribution, oversampling examples from the top decile.

percentile)

3. “i have a business competitor one in specific,,he has noticed that my business is doing well and has decided too the price of the products so i cannot get customers” (91st percentile)

Conversely, the least challenge-focused questions are clearly from businesses in less harrowing situations. These questions tend to focus on expansion and growth, as opposed to fundamental business challenges. It is plausible that the AI mentor is more effective at helping entrepreneurs tackle these problems.¹⁸ Some of the least challenge-focused questions include:

1. “i have ksh 20000 to expand my business which it would specialize on selling food stuffs. i would like to do this expansion in a new location. what should i consider first to maximize my profit?” (1st percentile)
2. “i have been operating a cereals shop for five years and it is doing very well lately for the last two years.which strategies can i use to expand the business to increase capacity to serve many customers and increase the products i’m selling?” (2nd percentile)
3. “i run a salon and barber shop in one room, recently customers have started to gain interest on massage after shave, this means there is need for expansion and more resources, kindly help me to unlock this potential.” (17th percentile)

We test for differences in the difficulty of the questions posed by low and high performers by regressing our measure of *challenge-focused* on our binarized measure of baseline firm performance, clustering standard errors at the level of the entrepreneur. We find that high performers ask questions that are less challenge-focused than those asked by low performers ($\delta = -0.25$ s.d., $p = 0.04$, 95% CI = [-0.49,-0.01]). On the whole, these results are robust to whether we use a continuous, rather than binarized measure of baseline firm performance, whether we focus on only above-median length questions or include hard-to-classify shorter queries, and whether we run our analysis at the level of the message or the entrepreneur (See Appendix J for details). Figure 6 visualizes the relationship between baseline firm performance and question difficulty using a binned scatter plot. There is a clear negative relationship between an entrepreneur’s baseline performance and how challenge-focused their questions are. Using a continuous measure of baseline performance, we find that a one s.d. increase in an entrepreneur’s baseline performance is associated with a nearly 0.5 s.d. decrease in how challenged-focused their questions are ($\delta = -0.48$ s.d., $p = 0.03$, 95% CI = [-0.91,-0.06]). In summary, whether by choice or by necessity, low-performing entrepreneurs

¹⁸Table A15 shows above-median-length example questions pulled from the 1th to 49th percentiles of the *challenge-focused* distribution, oversampling examples from the bottom decile.

in our sample asked the AI mentor for assistance with more challenging tasks than high performers.

This finding helps reconcile our inequality-increasing heterogeneous treatment effects with recent results that suggest generative AI reduces differences in productivity (Noy and Zhang, 2023; Peng et al., 2023; Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023). Even if the returns to AI advice on any given task for low performers are greater than or equal to the returns for high performers (as suggested by other studies), who in our setting low performers tend to seek assistance with more challenging tasks could lead to the heterogeneous treatment effects that we observe. Figure 7 illustrates how the business impact of AI advice within a task can be radically different than the aggregate impact of AI advice when entrepreneurs, firms, and workers endogenously end up with different tasks for which they require and/or seek AI assistance.

4 Discussion

Our findings highlight that in more open-ended and complex contexts, the productivity and performance implications of generative AI fundamentally depend on the tasks for which firms and entrepreneurs use AI tools. In contrast to existing studies that find positive effects of generative AI on productivity within specific business tasks (Noy and Zhang, 2023; Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023), we find that access to generative AI negatively impacted low-performing entrepreneurs and positively impacted high-performing entrepreneurs, resulting in a failure to reject the null hypothesis of no average effect of generative AI across our entire sample. Analysis of the interaction logs between treated participants and the AI mentor suggests that the treatment effect heterogeneity we observe may stem from low-performing entrepreneurs asking for advice on particularly challenging problems. While the AI appears to have generated well-structured advice in response to these difficult questions, our findings suggest that when low-performing entrepreneurs actually put that advice into action, the end result was performance declines relative to our control group.

Beyond this contribution, our study is—to our knowledge—the first randomized control trial testing the impact of generative AI on firms in developing economies. Many people

in developing economies face barriers to getting helpful personalized feedback, mentorship, and training (Chatterji et al., 2019; Dimitriadis and Koning, 2023; Björkegren, 2023). While researchers have raised concerns that LLMs may be less effective when used to address problems in developing economies due to the dominance of AI training data from the U.S. and other developed economies (Tao et al., 2023), our findings show that generative AI-based tools can serve as an efficient source of personalized feedback for many entrepreneurs in low- and middle-income countries. That being said, we also find evidence that generative AI can widen the gap between low- and high-performing businesses in these contexts. More broadly, our findings are consistent with the idea that data and AI have the potential to benefit the billions of people and millions of firms in developing countries (Aiken et al., 2022; Choi et al., 2023).

Our paper also contributes to a large literature focused on small business training (McKenzie, 2021). A recent meta-analysis shows that in-person training programs increase profits (sales) by 10% (5%), with programs that leverage mentorship, peer learning, psychosocial training and localization yielding profit (sales) increases of 15% (11%) (Campos et al., 2017; McKenzie, 2021; Bossuoy et al., 2022; Dimitriadis and Koning, 2022). Our point estimate of a 15% increase in business performance for high performers suggests that AI mentors already have the capability to advise some businesses as well as the human mentors evaluated in prior work. While the AI caused performance decreases for low-performing entrepreneurs, prior work on small business training has largely not reported heterogeneity by baseline performance, leaving open the possibility that human mentorship may actually have a similar pattern of effects as our AI mentor. Finally, while our overall average treatment is not statistically significant, the effect in our preferred specification is positive and represents a 3.2% to 4.4% increase in profits and revenue. While our study was not powered to detect effects of this size, this point estimate suggests an effect size that is roughly one-third the size of the effect of human-to-human training programs. Given the low cost of AI mentorship, future work should test the impact of AI using larger samples to better test if the overall impact of AI mentorship is positive but smaller than what our study can reliably detect.

Our study is not without limitations. First, the AI mentor that we evaluate is built on top of GPT-4 with only a relatively short system prompt and was distributed to entrepreneurs

along with only a 5-minute-long online training survey. It is possible that by fine-tuning the language model using a corpus of relevant information, by iterating further on the system prompt provided to the model, and/or by providing more training to the entrepreneurs interacting with the AI mentor, we might be able to further improve the positive impacts we observe among high performers and reduce (or reverse) the adverse effects that we observe among low performers. Second, although the five-month-long duration of our experiment (three months pre-treatment, two months post-treatment) is considerably longer than the overwhelming majority of prior experiments measuring the economic impacts of generative AI (cf. Brynjolfsson, Li, and Raymond, 2023), there may still be longer-term impacts of exposure to the AI mentor that we do not observe. Such longer-term impacts could be driven by a number of different factors, including learning among the entrepreneurs in our sample, and like the effects we measured after two months, may also be heterogeneous with respect to pre-treatment business performance.¹⁹ Finally, our experiment took place from May to November 2023, a period of time during which exposure to and experience with generative AI tools such as ChatGPT was still low, particularly in developing market contexts such as Kenya.²⁰ As both 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 (Raymond, 2023).

Our findings also suggest various opportunities for further research. As previously mentioned, the fact that our AI mentor may have had positive impacts for even a subset of entrepreneurs suggests that generative AI-based tools merit further study as a low-cost alternative to other interventions that have typically been studied in developing economies. One important piece of this research agenda is better understanding the mechanisms that drive entrepreneurs, firms, and workers to ask for AI assistance on some tasks but not others; while we find that low-performing entrepreneurs asked more challenging questions, we do not know if this was by choice, or by necessity. Similarly, we find that some entrepreneurs use

¹⁹In fact, learning that the AI mentor’s recommendations do not improve (and may even hurt) performance may, with time, lead some entrepreneurs to realize their best option is to pivot or shut down their businesses entirely (Koning, Hasan, and Chatterji, 2022).

²⁰While this is true in a general sense, it is worth noting that one-third of the entrepreneurs in our sample reported using ChatGPT at least once a week in our pre-treatment surveys.

AI despite the presence of negative returns, yet our analysis sheds little light on why this is the case. Relatedly, it's possible that the AI mentor's heterogeneous impacts are at least in part attributable to biases in GPT-4's training data, and that a model trained on better data could yield positive AI returns for even the most challenging business tasks (Tao et al., 2023). Finally, the AI mentor in our experiment does not learn, nor have we explored how the AI mentor impacted entrepreneurial learning. In theory, the data from our experiment could be used to improve the AI mentor's recommendations, opening up the possibility not only of better mentors, but also of strategic concerns around asking for advice from an AI tool that may also be used by a competitor (Brynjolfsson, Li, and Raymond, 2023). How entrepreneurs will change their ideas, skills, and judgment in the presence of generative AI tools also remains an open question (Lebovitz, Lifshitz-Assaf, and Levina, 2022; Gaessler and Piezunka, 2023; Boussioux et al., 2023; Girotra et al., 2023; Hui, Reshef, and Zhou, 2023).

While there is still much to learn, our field experiment demonstrates the value of studying the impact of AI within the actual day-to-day decisions and tasks that take place within businesses (Harrison and List, 2004). That our treatment effects differ from prior estimates on the impact of AI provides further evidence of the value of measuring the impact of interventions through "in vivo" experimentation, above and beyond measuring the impact of those same interventions through "in situ" experiments (Aral and Walker, 2012). Indeed, the findings from our field experiment highlight that not only are the impacts of AI uneven, but that they radically shift depending on the tasks that AI tools are used for.

References

- 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.
- Aiken, Emily, Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E Blumenstock. 2022. “Machine learning and phone data can improve targeting of humanitarian aid.” *Nature* 603 (7903):864–870.
- Anderson, Michael L. 2008. “Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American statistical Association* 103 (484):1481–1495.
- Anderson, Stephen J, Christy Lazicky, and Bilal Zia. 2021. “Measuring the unmeasured: Aggregating, anchoring, and adjusting to estimate small business performance.” *Journal of Development Economics* 151:102655.
- 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.
- Aral, Sinan and Dylan Walker. 2012. “Identifying influential and susceptible members of social networks.” *Science* 337 (6092):337–341.
- 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.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. “A multifaceted program causes lasting progress for the very poor: Evidence from six countries.” *Science* 348 (6236):1260799.
- Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. “Cross-country differences in productivity: The role of allocation and selection.” *American Economic Review* 103 (1):305–334.
- Baxter, Kathy M and Yoav Schlesinger. 2023. *How Companies Can Build Trustworthy AI Assistants*. Harvard Business Review Press.
- 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*.
- Bloom, Nicholas and John Van Reenen. 2007. "Measuring and explaining management practices across firms and countries." *The Quarterly Journal of Economics* 122 (4):1351–1408.
- Bossuroy, Thomas, Markus Goldstein, Bassirou Karimou, Dean Karlan, Harounan Kazianga, William Parienté, Patrick Premand, Catherine C Thomas, Christopher Udry, Julia Vaillant et al. 2022. "Tackling psychosocial and capital constraints to alleviate poverty." *Nature* 605 (7909):291–297.
- 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." Available at SSRN 4533642.
- Brooks, Wyatt, Kevin Donovan, and Terence R Johnson. 2018. "Mentors or teachers? Microenterprise training in Kenya." *American Economic Journal: Applied Economics* 10 (4):196–221.
- 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 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.
- Cai, Jing and Adam Szeidl. 2018. "Interfirm relationships and business performance." *The Quarterly Journal of Economics* 133 (3):1229–1282.
- 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*.
- 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.

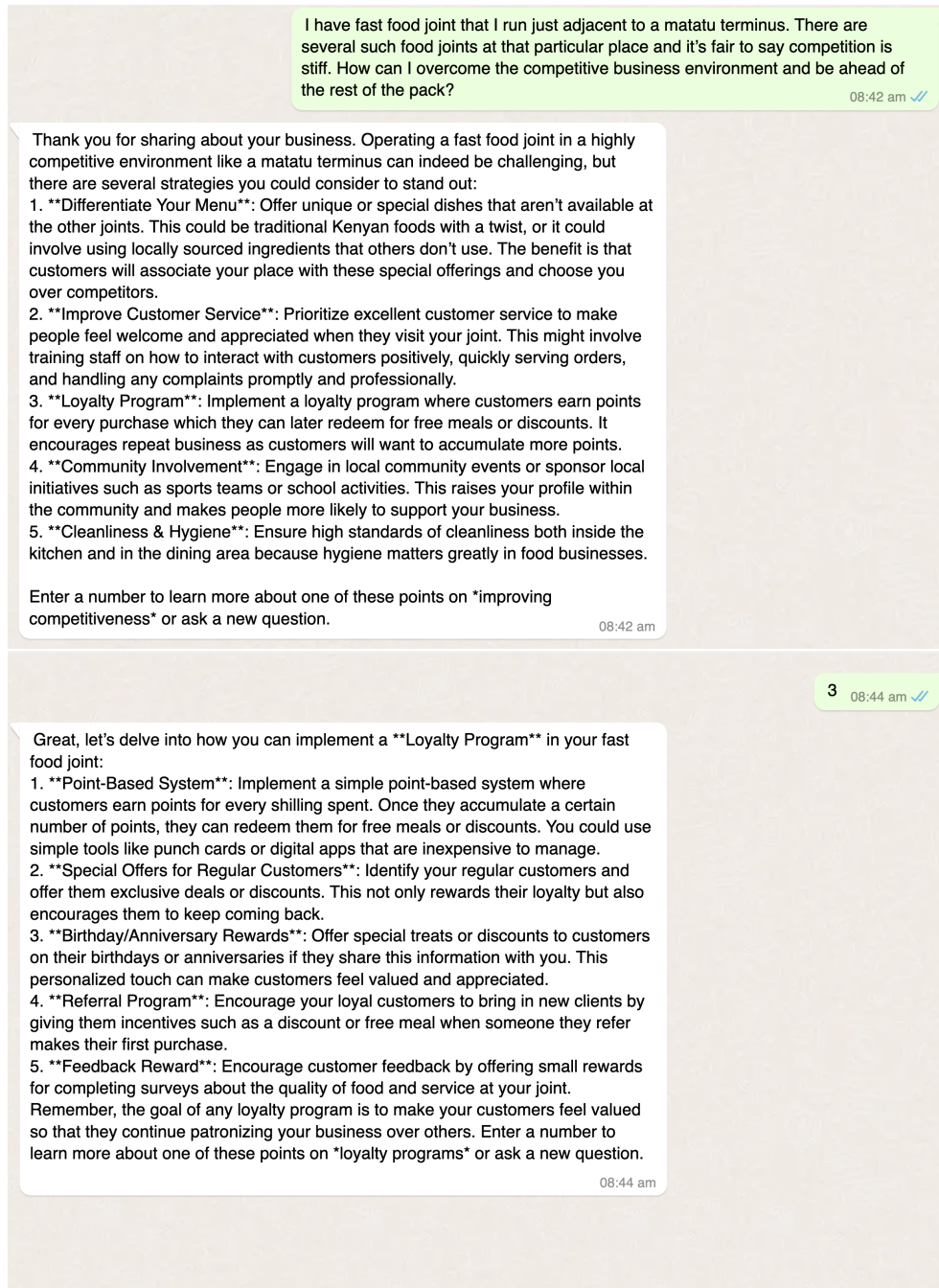
- Chen, Jiafeng and Jonathan Roth. 2023. “Logs with zeros? Some problems and solutions.” ArXiv: 2212.06080.
- Chen, Zenan and Jason Chan. 2023. “Large Language Model in Creative Work: The Role of Collaboration Modality and User Expertise.” SSRN 4575598.
- Choi, Jun Ho, Oliver Garrod, Paul Atherton, Andrew 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.
- Dalton, Patricio S, Julius Rüschenpöhler, Burak Uras, and Bilal Zia. 2021. “Curating local knowledge: Experimental evidence from small retailers in Indonesia.” *Journal of the European Economic Association* 19 (5):2622–2657.
- Davies, Elwyn, Peter Deffebach, Leonardo Iacovone, and David Mckenzie. 2024. “Training microentrepreneurs over Zoom: Experimental evidence from Mexico.” *Journal of Development Economics* 167:103244.
- Delecourt, Solène and Anne Fitzpatrick. 2021. “Childcare matters: Female business owners and the baby-profit gap.” *Management Science* 67 (7):4455–4474.
- Dell’Acqua, Fabrizio, Edward McFowland, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Kraymer, François Candelon, and Karim R Lakhani. 2023. “Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality.” SSRN 4573321.
- 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.
- . 2023. “Networking Frictions: Evidence from entrepreneurial networking events in Lomé.” Working Paper.
- 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. 2023. “Gpts are gpts: An early look at the labor market impact potential of large language models.” arXiv preprint arXiv:2303.10130.
- 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.
- Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff. 2012. “Using PDA consistency checks to increase the precision of profits and sales measurement in panels.” *Journal of Development Economics* 98 (1):51–57.

- Gaessler, Fabian and Henning Piezunka. 2023. "Training with AI: Evidence from chess computers." *Strategic Management Journal* .
- 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.
- 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.
- 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* .
- Hu, Krystal. 2023. "ChatGPT sets record for fastest-growing user base - analyst note — reuters.com." URL <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>. [Accessed 21-11-2023].
- 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.
- Imbens, Guido W and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.
- International Labor Organization. 2015. "SIYB Entrepreneur Level Materials." URL https://www.ilo.org/empent/areas/start-and-improve-your-business/WCMS_192062/lang--en/index.htm.
- 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.
- Kim, Hyunjin. 2023. "The value of competitor information: Evidence from a field experiment." Working paper.
- Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75 (1):83–119.
- Koning, Rembrand, Sharique Hasan, and Aaron Chatterji. 2022. "Experimentation and start-up performance: Evidence from A/B testing." *Management Science* 68 (9):6434–6453.
- Kozlowski, Austin C, Matt Taddy, and James A Evans. 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84 (5):905–949.
- 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.

- Lakens, Daniël. 2017. “Equivalence tests: A practical primer for t tests, correlations, and meta-analyses.” *Social psychological and personality science* 8 (4):355–362.
- 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.
- 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.
- 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.” Working Paper.
- 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.” .
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. “Distributed representations of words and phrases and their compositionality.” *Advances in Neural Information Processing Systems* 26.
- Moore, Don A. 2022. “Overprecision is a property of thinking systems.” *Psychological Review*.
- 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.”
- 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* .
- Raymond, Lindsey. 2023. “The Market Effects of Algorithms.” Working Paper.
- 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.
- Simonsohn, Uri. 2015. “Small telescopes: Detectability and the evaluation of replication results.” *Psychological science* 26 (5):559–569.

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

Figure 1: An example of a real interaction between a participant and the AI mentor, illustrating how we structured the AI mentor to push entrepreneurs to dig deeper and learn more when asking business questions.



Notes: This figure presents a rendering of an entrepreneur-AI interaction in WhatsApp. The text is original (including the typos). A “matatu” is a word commonly used to describe a minibus in Kenya. In Appendix H, we display the full conversation histories for two treated entrepreneurs. In Table A14 and Table A15 we provide additional example questions illustrating variations in how challenging the business tasks described in the questions are.

Figure 2: Overview of the experimental timeline from the date of the first survey

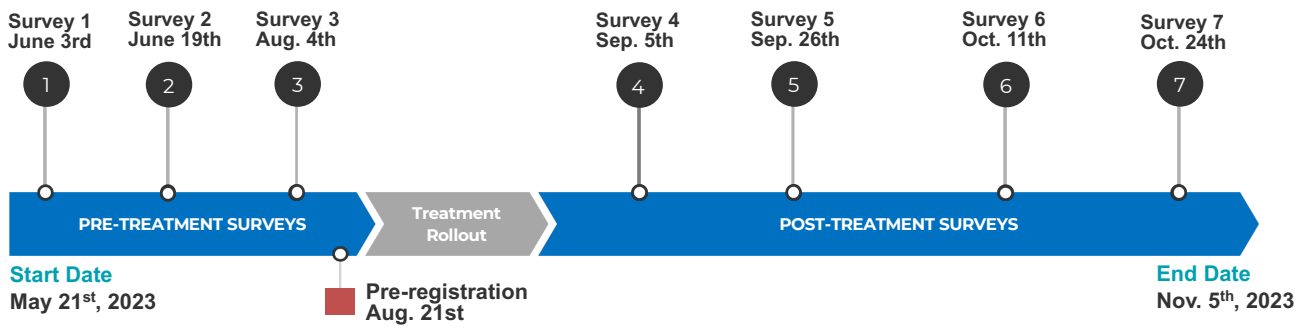
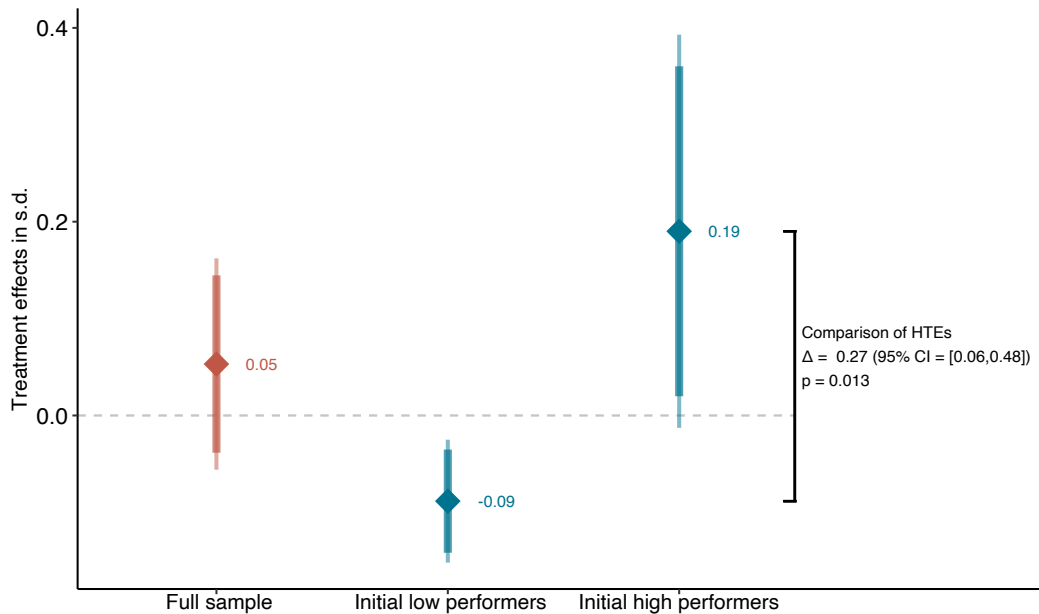
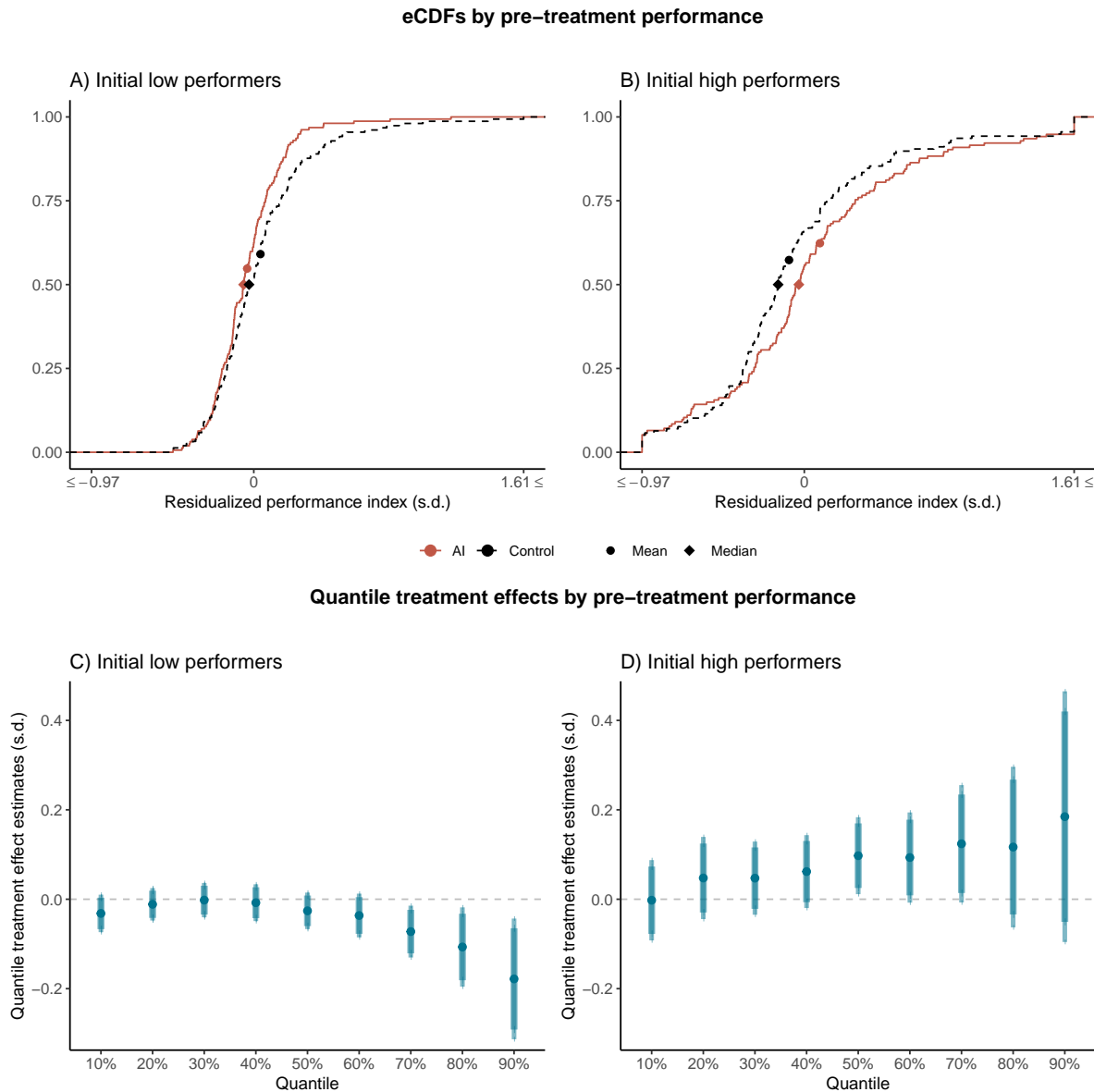


Figure 3: Treatment effect estimates reveal an overall null effect for our full sample of entrepreneurs, but a negative treatment effect for low performers (defined as having below-median pre-treatment performance), and a positive treatment effect for high performers (defined as having above-median pre-treatment performance).



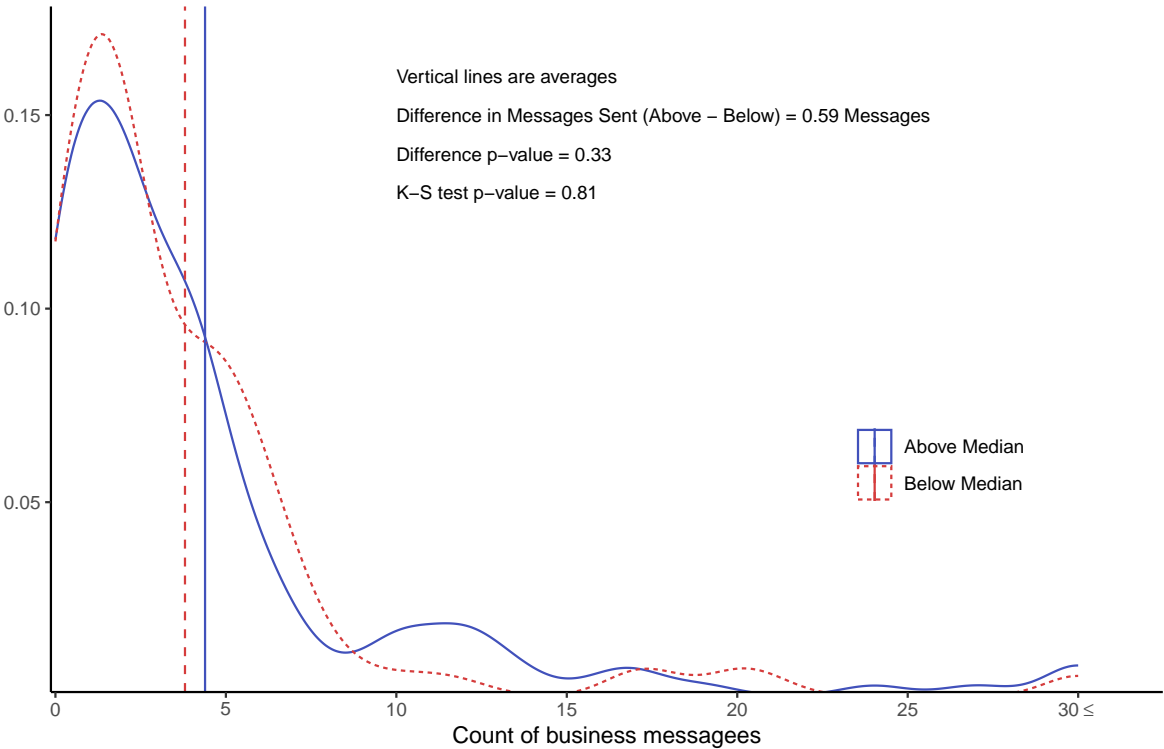
Notes: The effect of access to the AI mentor on our index of business performance winsorized at the 99% level, restricting to participants who have completed the panel. Outcomes are presented in standard deviations. Effects are estimated using a regression model that controls for pre-treatment outcome levels and baseline covariates selected using a double-LASSO procedure. Controls selected by the double-LASSO are report in Table A4. Error bars present 90 and 95% confidence intervals. The presented p -value is from a two-tailed test of the difference in treatment effects between entrepreneurs with below versus above median pre-treatment performance. See Table A12 for p -values corrected for multiple inferences following Benjamini and Hochberg (1995)

Figure 4: The distribution of changes in performance shifts to the left for treated low performers, whereas it shifts to the right for treated high performers, and these shifts are reflected in estimated quantile treatment effects.



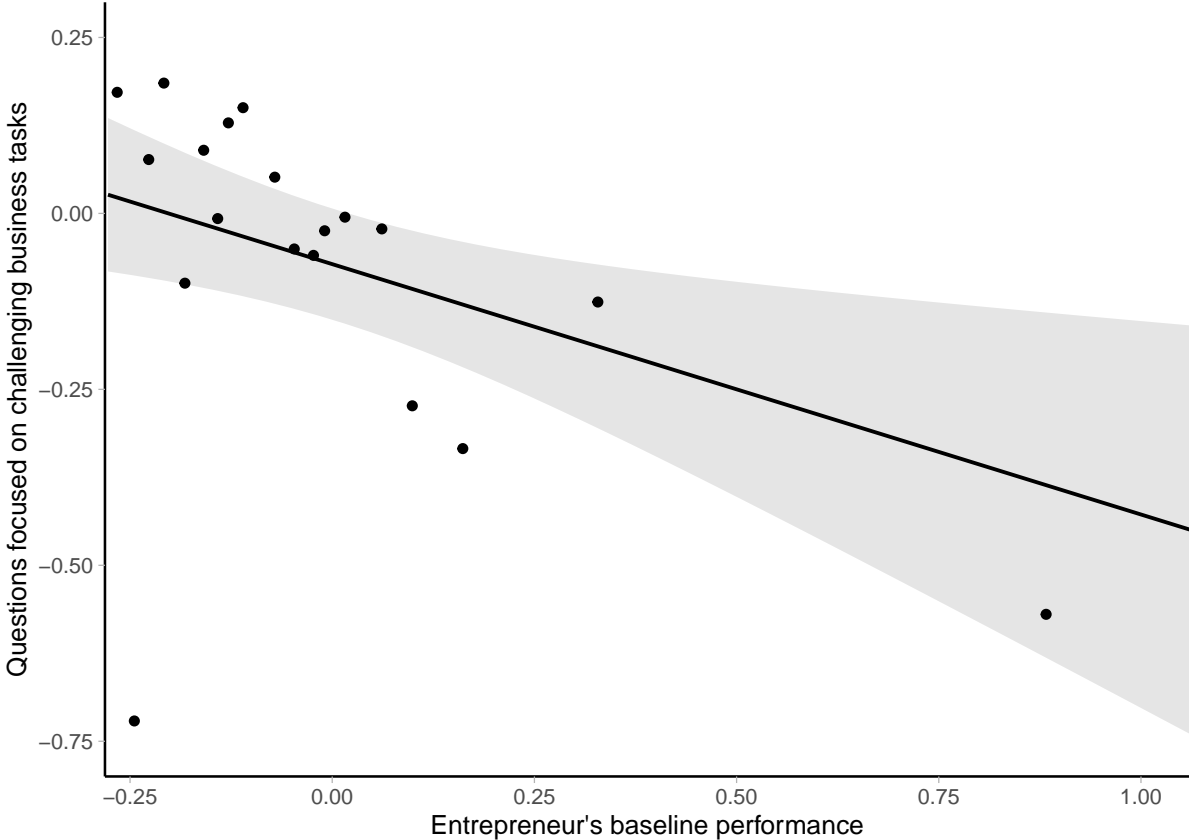
Notes: Panels A and B present cumulative density plots of the standardized index of business performance winsorized at the 99% level. It compares those with initial low performance (below-median pre-treatment performance; left panel) to those with initial high performance (above-median pre-treatment performance; right panel), across both treatment and control groups. The performance index is residualized using the double-LASSO-selected covariates reported in Table A4 and for stratification block and time period. Panels C and D present quantile treatment effect estimates for initial low and initial high performers at each decile, again residualizing using the same covariates as in the eCDFs. Error bars present 90% and 95% bootstrapped confidence intervals clustered at the individual level. All results are restricted to participants who have completed the panel.

Figure 5: The distribution of engagement with our AI mentor is the same for initially high- and low-performing entrepreneurs.



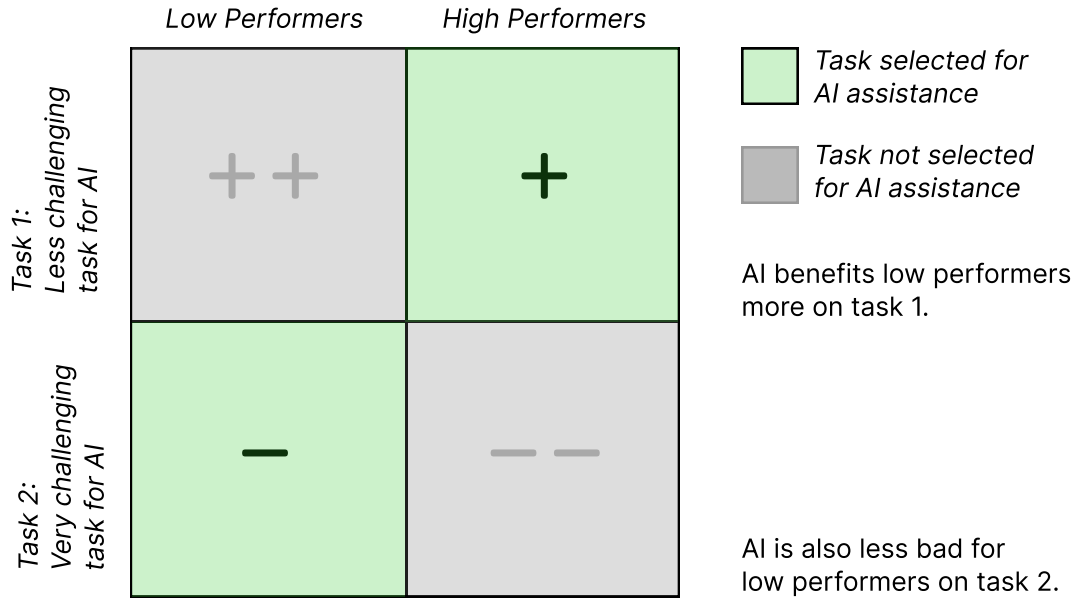
Notes: Kernel density plot showing the distributions of the number of business messages sent to the AI mentor for below-median (red, dashed line) participants versus above-median (blue, solid line) participants, winsorized at the 99% level. Vertical lines denote the average number of messages sent.

Figure 6: Entrepreneurs with stronger baseline performance asked the AI mentor about less challenging business tasks, whereas entrepreneurs with worse initial performance asked about more challenging tasks.



Notes: This binned scatterplot shows the relationship between how *challenge-focused* a question is (y-axis) and the entrepreneur’s baseline pre-treatment level of performance (x-axis). Each dot represents the average of our *challenge-focused* variable for each question for each ventile of the baseline performance distribution. The units for both axes are standard deviations. The plot only includes entrepreneurs in our sample who asked at least one business question of the AI. The black line corresponds to the regression estimate in Model 4 of Table A16 and gray-shaded region shows the 95% confidence intervals.

Figure 7: Returns to AI assistance by task type and baseline pre-AI performance.



Low performers, despite having better within task AI returns, have worse AI returns than high performers due to task selection.

Notes: Two-by-two illustrating how even if low performers benefit more than high performers from AI on any given task, if low performers are more likely to need AI assistance on more challenging tasks, the end result is that AI will lead to greater returns for high performers than low performers. The values in each cell represent the returns to AI use with ++ indicating positive and large returns, + positive but somewhat smaller returns, - - very negative returns, and - negative but less so. Figure A11 presents a more detailed numerical example that illustrates how task selection can yield the estimates we present in Figure 3 even if low performers benefit more from AI on each task, as suggested by recent work (e.g., Noy and Zhang, 2023; Peng et al., 2023; Brynjolfsson, Li, and Raymond, 2023; Dell’Acqua et al., 2023).

Online Appendix

Table of Contents

A Appendix Figures	33
B Appendix Tables	45
C Experimental Design and Timeline	65
D Detailed Intervention Overview	69
E Pre-analysis Plan	76
E.1 Overview of Pre-registration	76
E.2 Econometric Strategy	77
E.3 Overview of Measures	78
E.4 Explanation of Preferred Functional Forms	82
E.5 Hypothesis Testing and Multiple Comparisons	83
F Ethical Considerations	84
G Accounting for Spillovers	86
H Detailed Case Examples	88
I Text Classification and Coding	102
J Word Embedding Analysis of Question Challenge Focus	107

A Appendix Figures

Figure A1: Recruitment ad



Figure A2: The distribution of study participants across Kenya

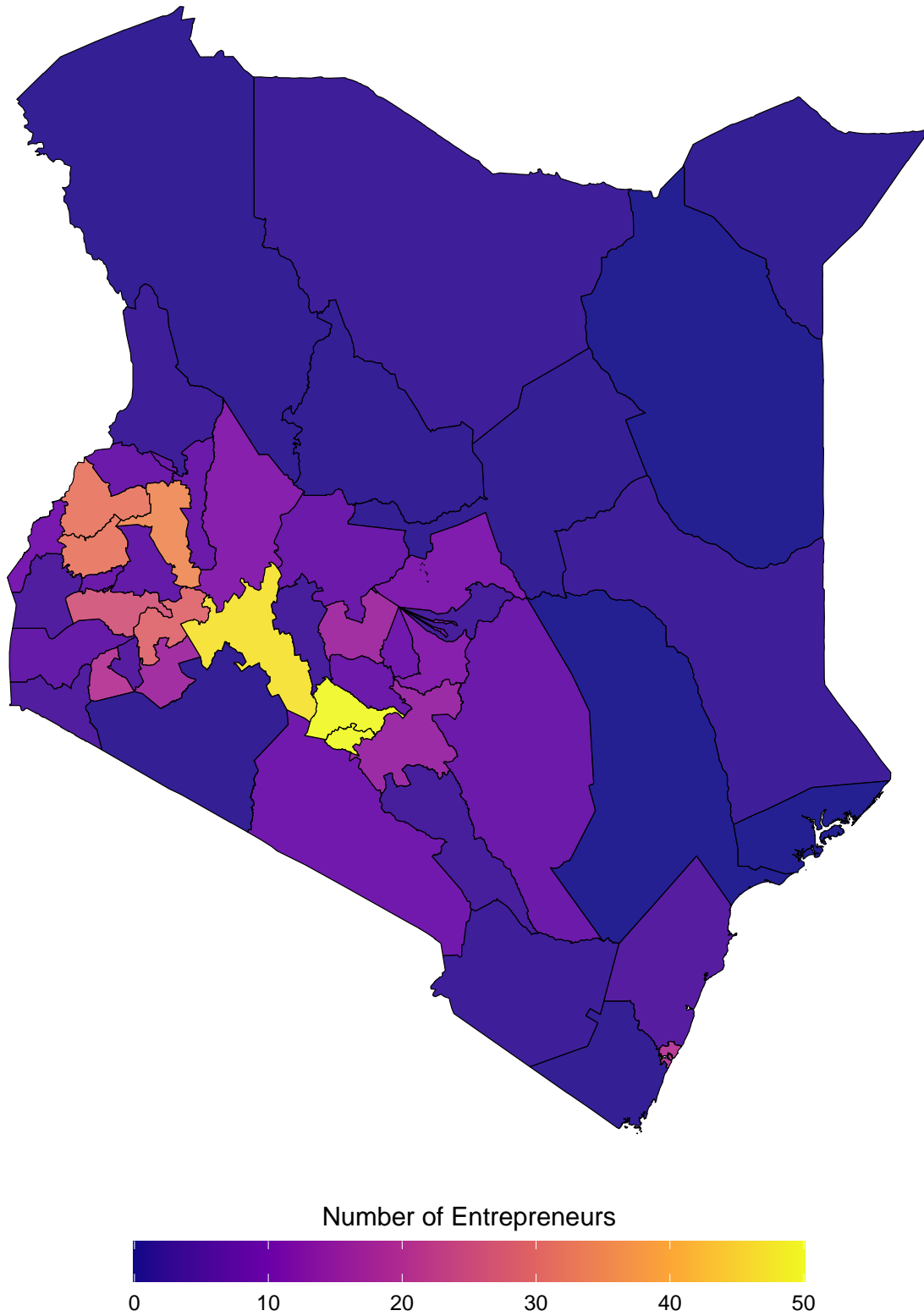
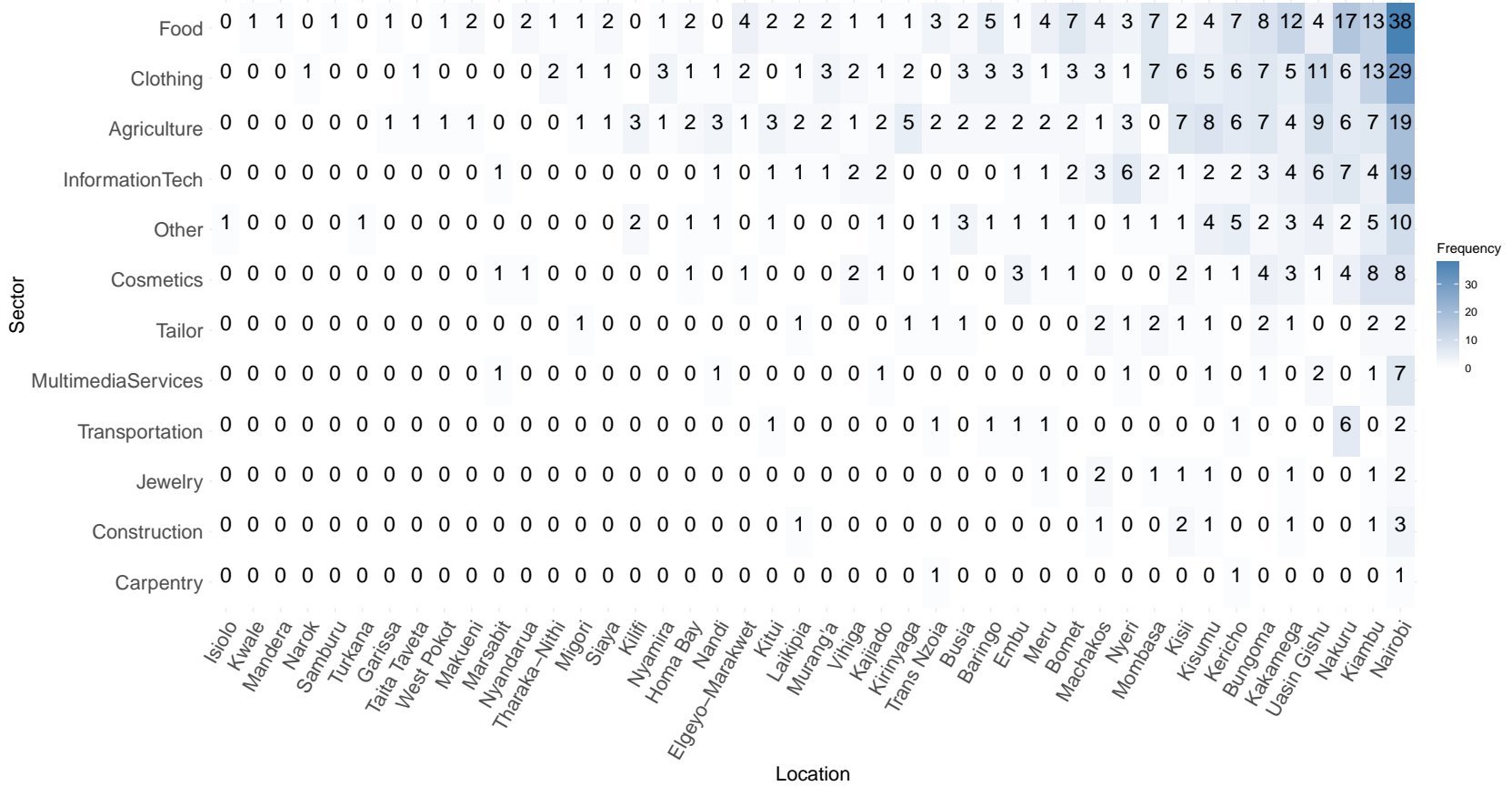


Figure A3: Distribution of Sectors and Kenyan Counties

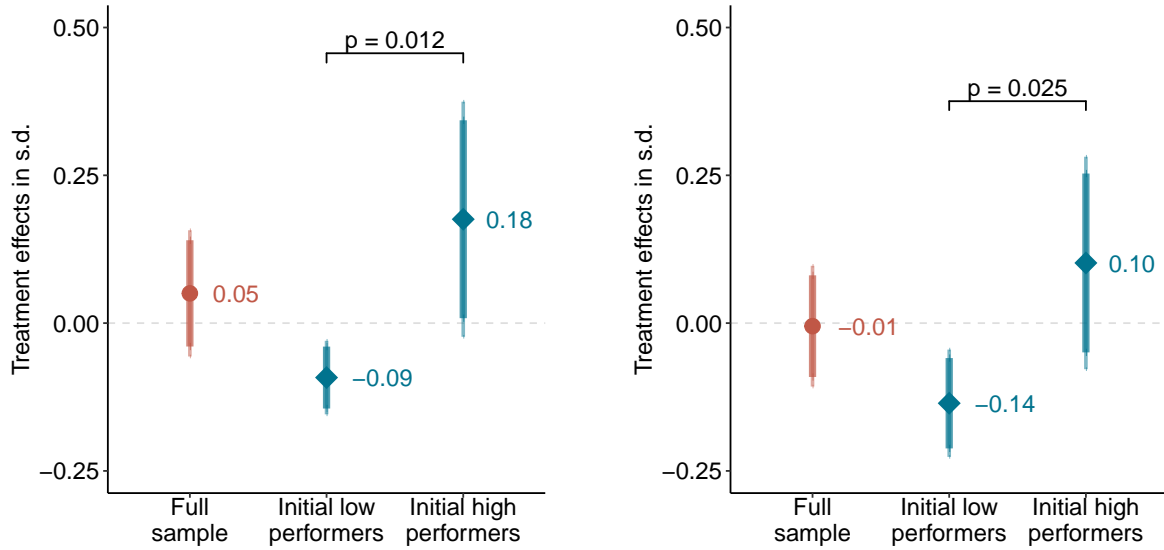


Notes: This figure depicts the distribution of Kenyan counties and firm sectors. Note that each participant may be involved in multiple sectors.

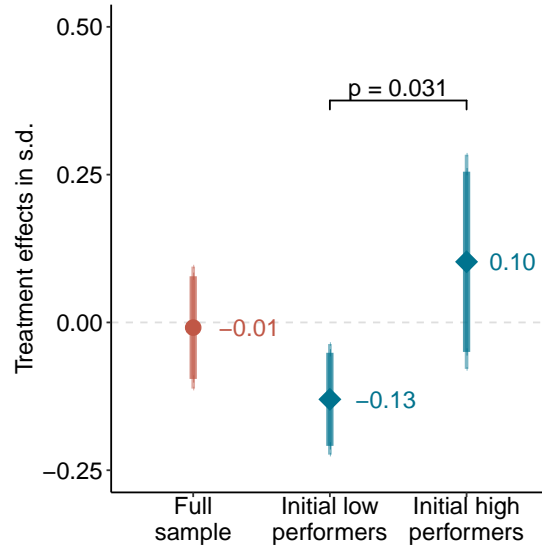
Figure A4: Performance effects (Robustness checks)

(A) Winsorized at the 99% level, all observations

(B) Winsorized at the 95% level, all observations

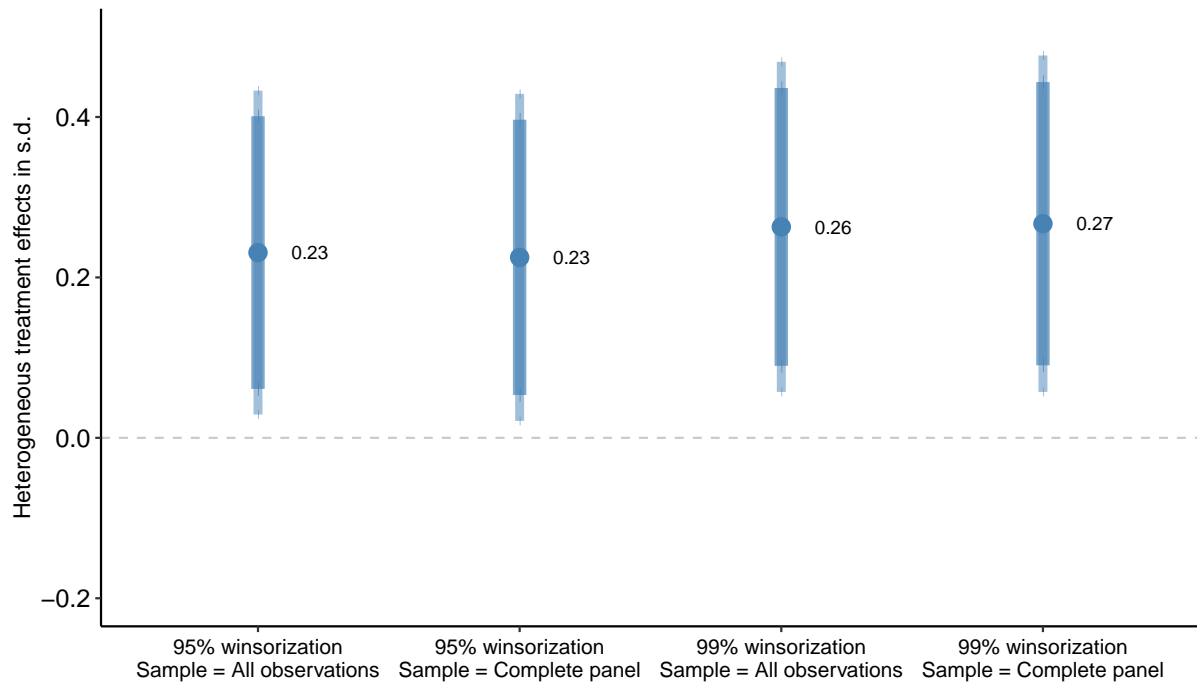


(C) Winsorized at the 95% level, completed panel



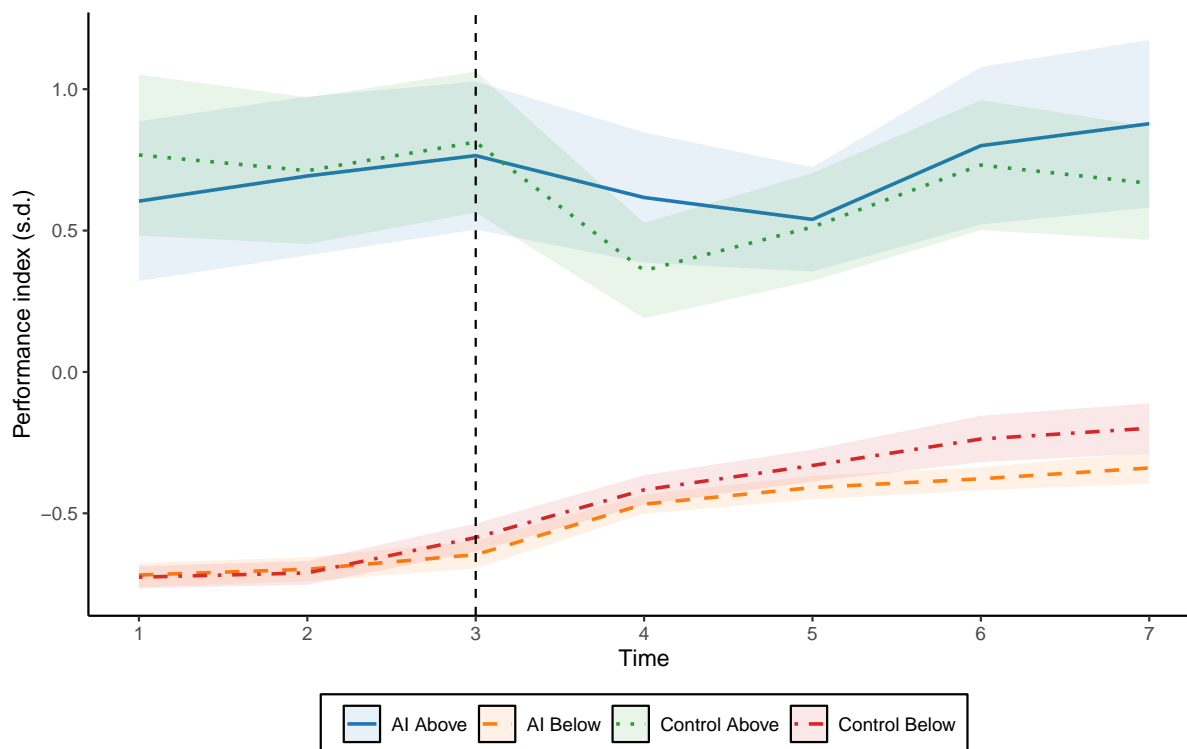
Notes: This figure presents the causal effect of access to the AI mentor on business performance. Outcomes are presented in standard deviations. Panel (A) and (B) present results winsorized at the 99% and 95% respectively for all participants' observations. Panel (C) presents results winsorized at the 95% level restricting to individuals who have completed the panel. Effects are estimated using an intent-to-treat specification that controls for pre-treatment outcome levels and baseline covariates selected using a double LASSO procedure. The presented p -values are from a two-tailed test of the difference in treatment effects between entrepreneurs with below versus above median pre-treatment performance. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Error bars present 95 and 90% confidence intervals.

Figure A5: Heterogeneous treatment effect estimates for performance specifications



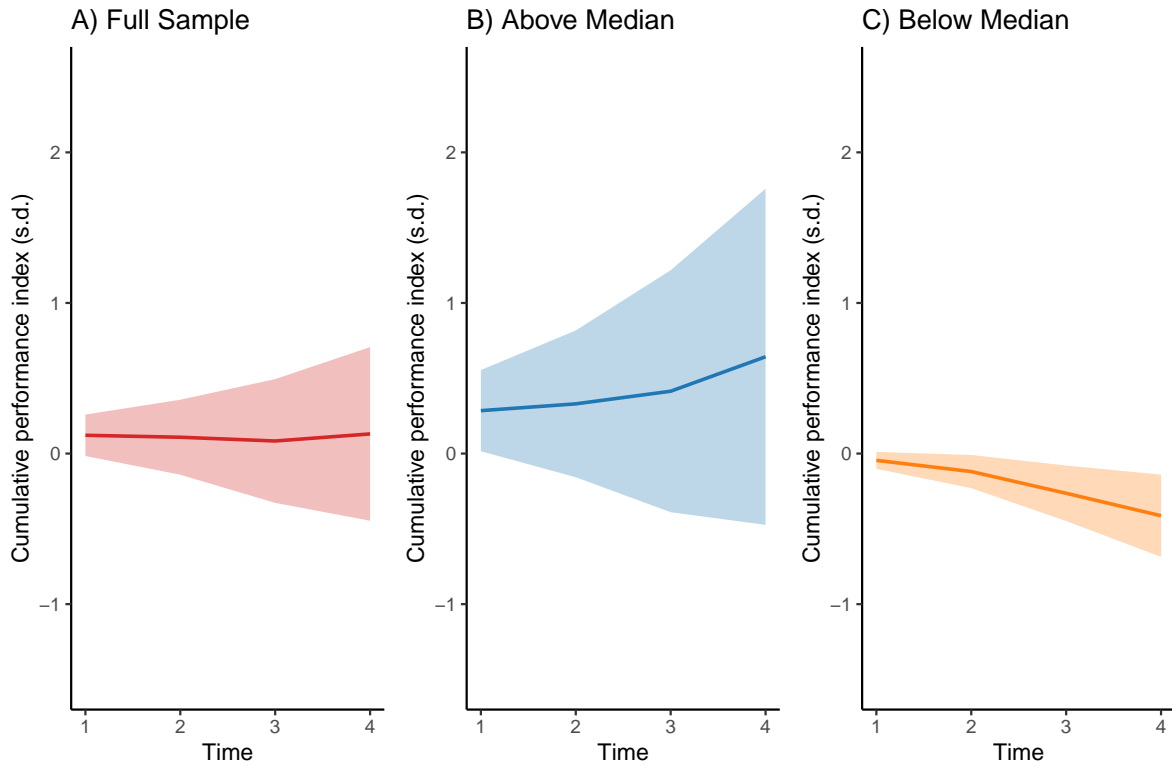
Notes: This figure presents heterogeneous treatment effect estimates for entrepreneurs with above (versus below) median pre-treatment performance. Outcomes are presented in standard deviations. Each point represents a different winsorization/sample robustness check. Effects are estimated using an intent-to-treat specification that controls for pre-treatment outcome levels and baseline covariates selected using a double LASSO procedure. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Error bars present 95 and 90% confidence intervals clustered at the individual level.

Figure A6: Performance time trends



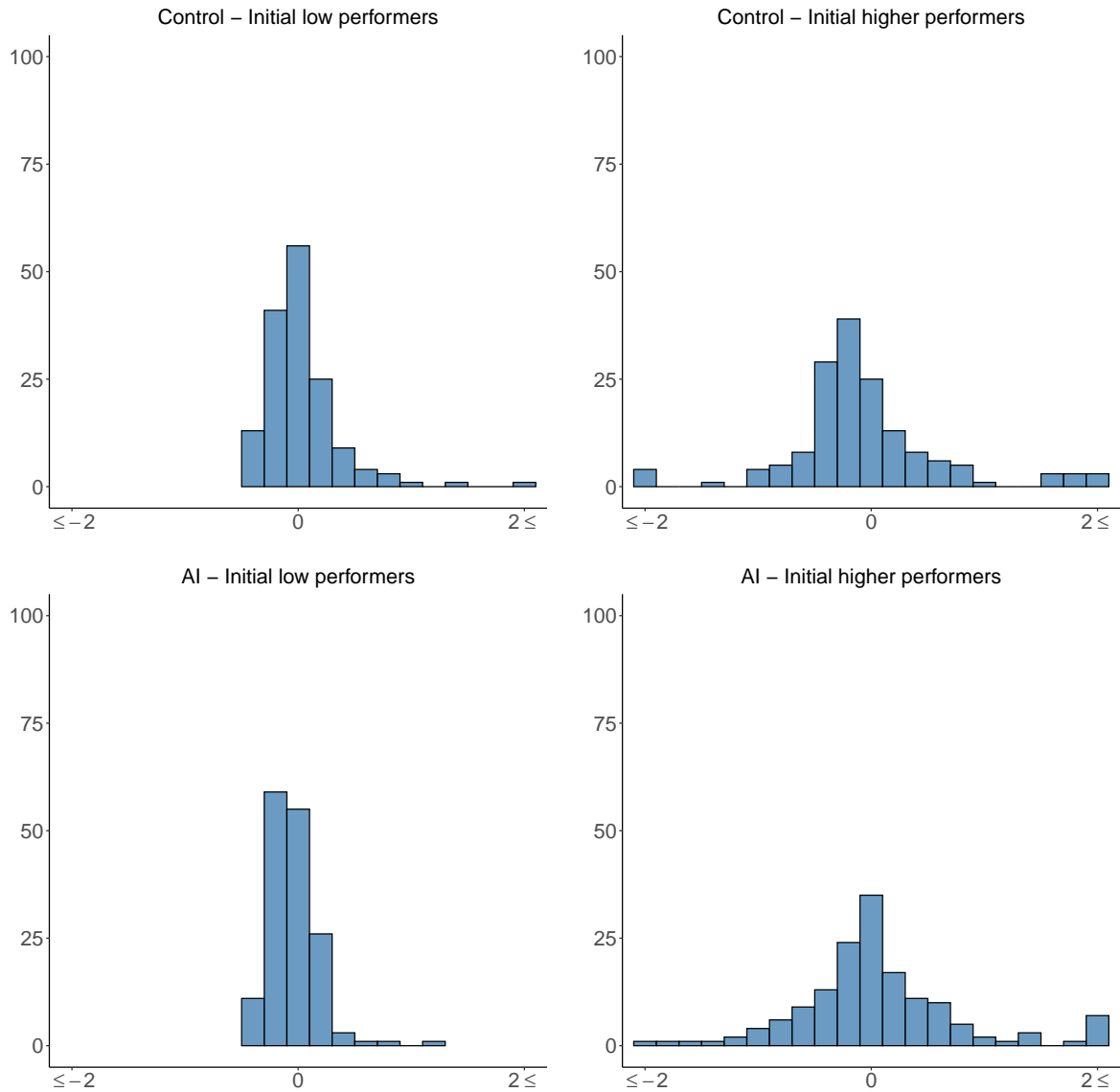
Notes: This figure presents time trends for the average performance index for above and below-median pre-treatment entrepreneurs who were randomly assigned to receive access to the AI mentor. Error bars display 95% confidence intervals around the mean for each group at each time-point. Treatment was administered following the third time-point.

Figure A7: Cumulative effects on performance time trends



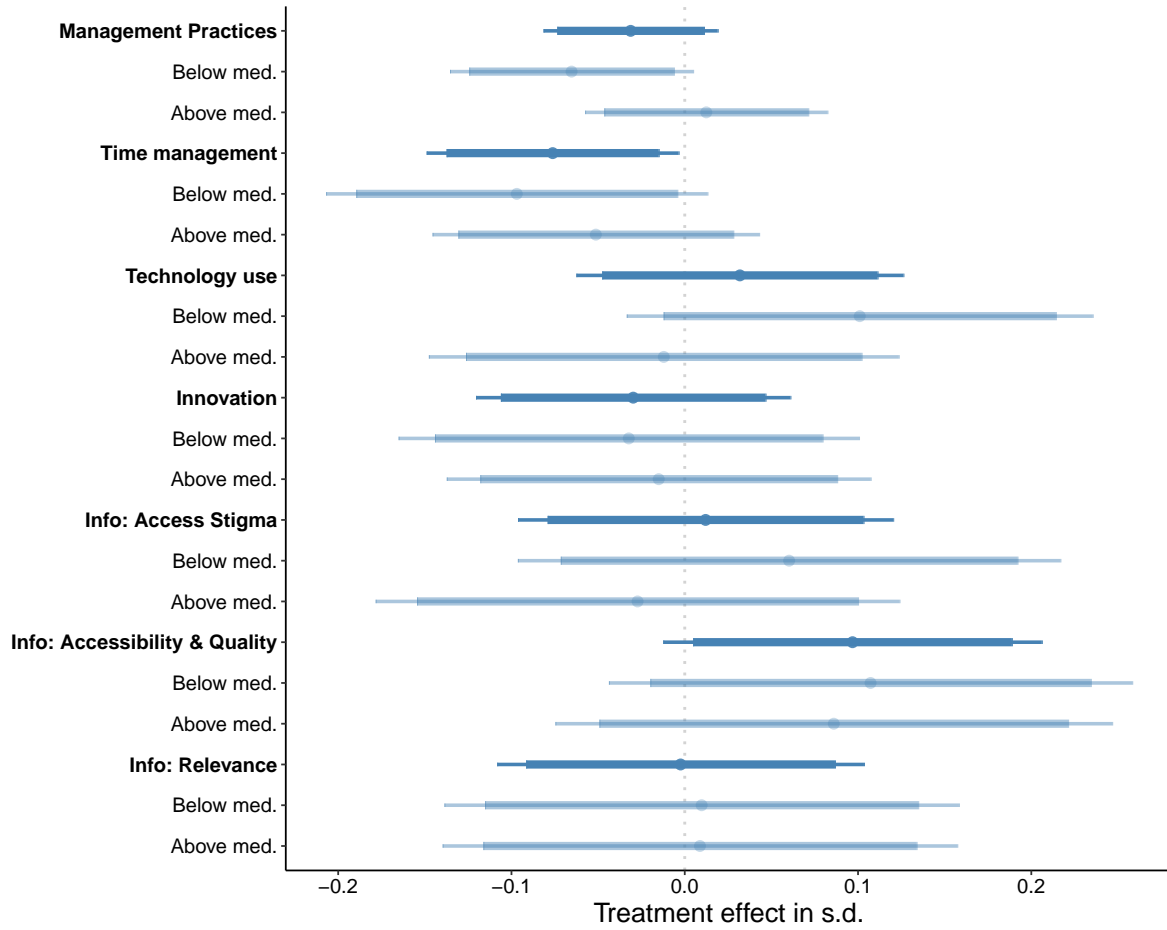
Notes: This figure presents cumulative time trends of treatment effects. Panel A depicts cumulative effects across all participants. Panels B and C present cumulative treatment effects for Above and Below-median participants respectively. Error bars display 95% confidence intervals around the mean for each group in each post-treatment time period.

Figure A8: Residualized Performance Indices by Condition and Pre-treatment Performance



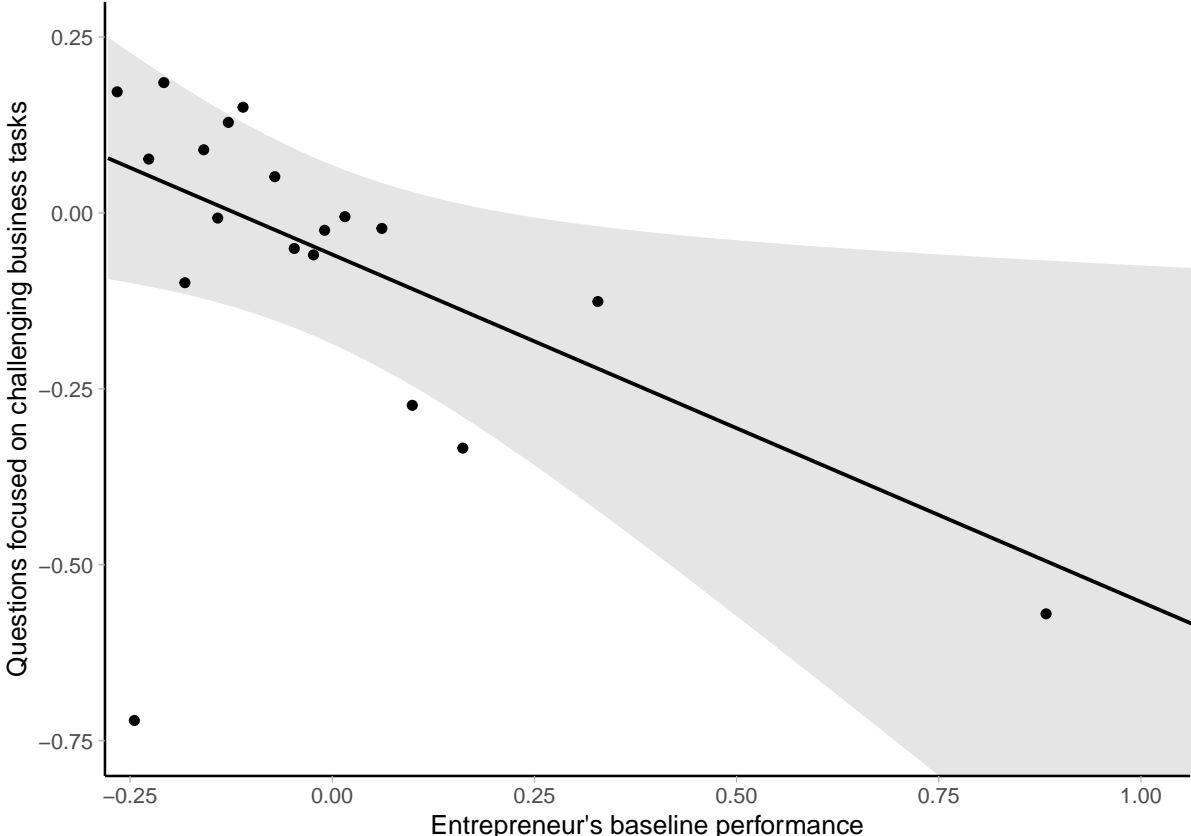
Notes: This figure presents histograms of the standardized performance index by experimental group and whether the entrepreneur was above or below median for pre-treatment performance. The index is residualized using the double-LASSO-selected covariates reported in Table A4 and for stratification block and time period. Results are restricted to participants who have completed the panel, and observations are averaged at the entrepreneur.

Figure A9: Effects on mechanism questions



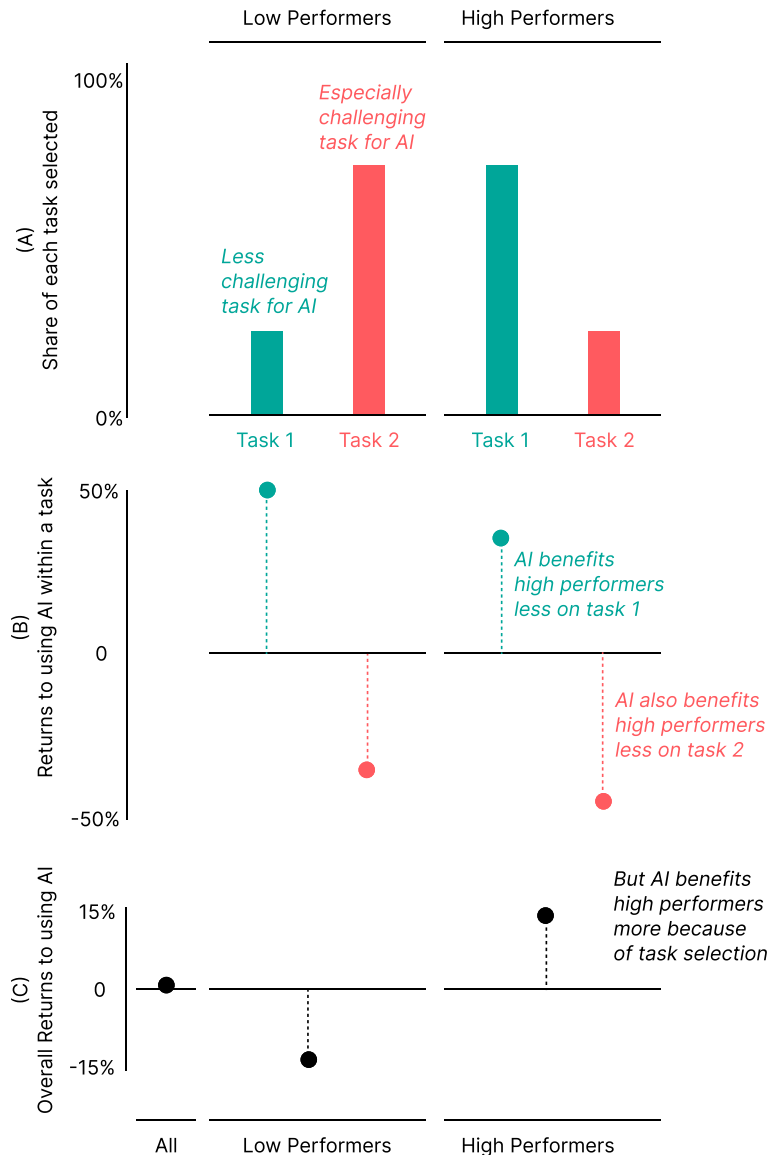
Notes: This figure depicts treatment effect estimates for our seven pre-registered mechanism indices. Dark bars present average treatment effect estimates with lasso covariates selection among all responding participants. Light bars present treatment effect estimates for above- and below-median initially performing firms using the 99% winsorized pre-treatment performance index to generate the median split. Error bars present 95 and 90% confidence intervals. See Table A12 for p -values corrected for multiple inferences following Benjamini and Hochberg (1995)

Figure A10: This figure is conceptually the same as Figure 6, but the data is plotted at the entrepreneur-level rather than the message-level. We again see that entrepreneurs with stronger baseline performance ask the AI mentor about less challenging business tasks.



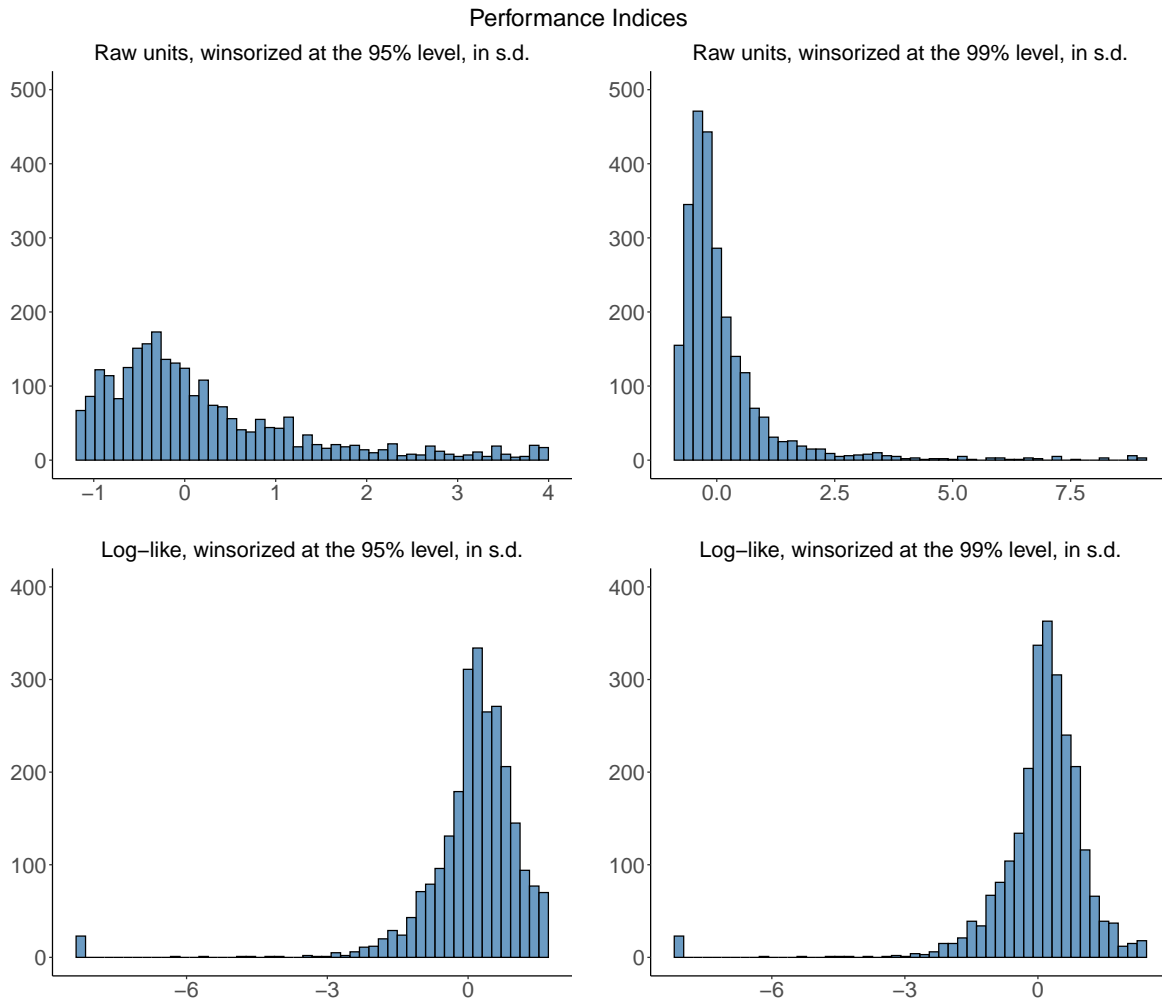
Notes: This binned scatterplot shows the relationship between how *challenge-focused* an entrepreneur's questions were (y-axis) and the entrepreneur's baseline pre-treatment level of performance (x-axis). Each dot represents the average of our *challenge-focused* variable for each ventile (one-twentieth) of the baseline performance distribution. The units for both axes are standard deviations. The plot only includes entrepreneurs in our sample who asked at least one business question to the AI mentor. The black line corresponds to the regression estimate in column 3 of Table A16 and gray-shaded region shows the 95% confidence intervals.

Figure A11: Numerical example showing that even if within tasks, low performers benefit more than high performers, differences in task selection can drive performance effects in the opposite direction.



Notes: The three panels above present a stylized example illustrating how task selection can radically shift the distributional consequences of AI. In Panel (A) we assume 75% of low performers select challenging tasks (Task 1) for AI assistance compared to 25% of high performers. In Panel (B) we assume the treatment effects (returns) of using AI for task 1 are 50% for low performers as against only 35% for high performers; for task 2 we assume -35% for low and -45% for high performers. Yet, in Panel (C) we see the overall impact of AI mirrors our experimental findings: despite assuming the returns to AI assistance are greater for low performers on both tasks, the overall returns from AI are essentially null at 1.25%, the returns for high performers are 15%, and the returns for low performers are -13.75%.

Figure A12: Distribution of outcomes



Notes: These figures present the distribution of four functional forms of our performance index. The top two figures depict results measured in raw units (KSH), 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 99% level. In all cases, these results are presented in standard deviations for comparability.

B Appendix Tables

Table A1: Balance

	Full analytic sample				Any post-treatment				Completed Panel			
	All	AI	Control	<i>p</i>	All	AI	Control	<i>p</i>	All	AI	Control	<i>p</i>
Business features												
Pre-treatment performance												
99pct winsorize (s.d.)	-0.00	-0.02	0.02	0.58	0.00	-0.02	0.02	0.61	0.01	-0.01	0.03	0.64
95pct winsorize (s.d.)	0.00	-0.03	0.04	0.34	0.00	-0.03	0.03	0.38	0.01	-0.02	0.04	0.39
Log-like 99pct winsorize (s.d.)	-0.00	-0.02	0.02	0.53	-0.00	-0.02	0.02	0.55	0.01	-0.02	0.03	0.55
Log-like 95pct winsorize (s.d.)	-0.00	-0.02	0.02	0.53	-0.00	-0.02	0.02	0.56	0.01	-0.02	0.03	0.56
Other business features												
Management practices count	13.10	13.08	13.12	0.88	13.11	13.08	13.13	0.84	13.12	13.11	13.14	0.87
Oldest bus.: less than 1 year	0.20	0.20	0.21	0.86	0.21	0.20	0.21	0.78	0.20	0.20	0.21	0.84
Oldest bus.: 1 year	0.41	0.41	0.40	0.79	0.41	0.42	0.40	0.73	0.41	0.42	0.40	0.68
Oldest bus.: 2 years	0.23	0.22	0.24	0.71	0.23	0.22	0.23	0.64	0.23	0.22	0.24	0.57
Oldest bus.: 3 or more years	0.16	0.16	0.15	0.80	0.16	0.16	0.15	0.71	0.16	0.16	0.15	0.74
Sector												
Agriculture	0.19	0.18	0.19	0.78	0.19	0.18	0.19	0.78	0.19	0.19	0.19	0.92
Cosmetics	0.07	0.06	0.08	0.26	0.07	0.06	0.08	0.26	0.07	0.05	0.08	0.16
Carpentry	0.00	0.01	0.00	0.57	0.00	0.01	0.00	0.57	0.00	0.01	0.00	0.56
Information technology	0.11	0.14	0.09	0.05	0.11	0.13	0.09	0.09	0.11	0.14	0.09	0.08
Jewelry	0.02	0.02	0.01	0.21	0.02	0.02	0.01	0.21	0.02	0.02	0.01	0.20
Multimedia services	0.03	0.03	0.02	0.13	0.03	0.03	0.02	0.13	0.03	0.04	0.02	0.13
Clothing	0.21	0.22	0.20	0.49	0.21	0.22	0.19	0.37	0.21	0.23	0.19	0.32
Food	0.27	0.24	0.30	0.09	0.27	0.24	0.30	0.11	0.27	0.24	0.30	0.09
Tailor	0.03	0.03	0.03	0.84	0.03	0.03	0.03	0.98	0.03	0.03	0.03	0.81
Transportation	0.02	0.02	0.02	0.98	0.02	0.02	0.02	0.98	0.02	0.02	0.02	1.00
Other	0.08	0.09	0.08	0.60	0.09	0.09	0.08	0.60	0.09	0.09	0.08	0.67
Participant features												
Female	0.32	0.31	0.33	0.48	0.32	0.31	0.33	0.48	0.31	0.30	0.33	0.39
Children	0.73	0.71	0.76	0.48	0.73	0.70	0.76	0.38	0.73	0.69	0.76	0.35
Age	25.86	25.73	25.99	0.54	25.85	25.71	25.99	0.51	25.86	25.72	26.01	0.50
Complete college	0.53	0.53	0.53	0.82	0.53	0.54	0.52	0.76	0.53	0.53	0.53	0.87
Started college	0.37	0.37	0.37	0.90	0.37	0.37	0.37	0.90	0.37	0.38	0.37	0.87
No college	0.10	0.09	0.11	0.56	0.10	0.09	0.11	0.47	0.10	0.09	0.10	0.59
ChatGPT at least once a week	0.36	0.34	0.38	0.31	0.36	0.34	0.38	0.27	0.36	0.34	0.38	0.28
Agreeableness	9.57	9.46	9.68	0.09	9.58	9.47	9.69	0.10	9.59	9.50	9.69	0.15
Conscientiousness	10.21	10.15	10.27	0.28	10.22	10.15	10.29	0.21	10.22	10.14	10.29	0.20
Extraversion	8.45	8.31	8.60	0.07	8.46	8.32	8.59	0.09	8.46	8.31	8.60	0.07
Openness	10.20	10.13	10.27	0.23	10.19	10.13	10.26	0.24	10.21	10.15	10.27	0.29
Neuroticism	7.26	7.31	7.22	0.51	7.26	7.31	7.22	0.55	7.27	7.31	7.22	0.55
Location (county)												
Machakos	0.02	0.02	0.02	0.81	0.02	0.03	0.02	0.81	0.02	0.03	0.02	0.79
Bomet	0.03	0.03	0.02	0.63	0.03	0.03	0.02	0.63	0.03	0.03	0.02	0.61
Nyeri	0.03	0.02	0.03	0.98	0.02	0.02	0.03	0.78	0.02	0.02	0.02	1.00
Kisii	0.03	0.02	0.04	0.07	0.03	0.02	0.04	0.07	0.03	0.02	0.05	0.07
Mombasa	0.03	0.03	0.03	0.67	0.03	0.03	0.03	0.67	0.03	0.04	0.03	0.65
Kisumu	0.04	0.03	0.05	0.22	0.04	0.03	0.05	0.22	0.04	0.03	0.05	0.23
Kericho	0.05	0.04	0.05	0.82	0.04	0.04	0.04	0.97	0.05	0.05	0.05	1.00
Bungoma	0.05	0.06	0.04	0.49	0.05	0.06	0.04	0.49	0.05	0.05	0.05	0.58
Kakamega	0.05	0.06	0.04	0.49	0.05	0.06	0.04	0.49	0.05	0.06	0.05	0.47
Uasin Gishu	0.05	0.05	0.06	0.83	0.06	0.05	0.06	0.83	0.05	0.05	0.06	0.72
Nakuru	0.07	0.07	0.08	0.62	0.07	0.07	0.08	0.73	0.07	0.07	0.08	0.76
Kiambu	0.08	0.07	0.09	0.45	0.08	0.08	0.09	0.45	0.08	0.07	0.09	0.47
Nairobi	0.21	0.21	0.21	0.99	0.21	0.21	0.21	0.99	0.21	0.20	0.21	0.84
Other	0.26	0.28	0.24	0.24	0.26	0.28	0.24	0.28	0.26	0.28	0.24	0.23
Completed Panel	0.97	0.97	0.98	0.35	0.98	0.97	0.99	0.25	1.00	1.00	1.00	0.35
Sample size	640	322	318		634	319	315		622	311	311	

Mean values and balance statistics for all participants all in the analytic sample (i.e., all treated participants), all participants who completed at least one post-treatment outcome survey, and all participants who completed the entire panel.

Table A2: Overview of measures

Outcomes (1)	Waves (2)	pre-registered (3)
Performance Index	Surveys 1-7	✓
Management Practices	Surveys 3-7	✓
Time Management	Surveys 4-7	✓
Technology Use	Surveys 4-7	✓
Innovation Activities	Surveys 4-7	✓
Info: Access Stigma	Surveys 4-7	✓
Info: Relevance	Surveys 4-7	✓
Info: Accessibility and Quality	Surveys 4-7	✓
Text Analysis	Post-treatment (~4-7)	✗

This table reports which outcomes were collected in each survey wave. Surveys 1-3 are pre-treatment waves and surveys 4-7 are post-treatment waves. The final column indicates whether the outcome was pre-registered.

Table A3: Effects on performance

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.035 (0.072)	0.050 (0.054)	-0.098** (0.034)	-0.082* (0.034)	-0.098** (0.034)	-0.092** (0.032)	0.163 (0.141)	0.176† (0.101)
AI mentor*(\geq median)			0.263† (0.145)	0.263* (0.105)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.041 (0.058)	-0.005 (0.052)	-0.147** (0.050)	-0.126** (0.047)	-0.147** (0.050)	-0.136** (0.046)	0.052 (0.104)	0.102 (0.092)
AI mentor*(\geq median)			0.198† (0.115)	0.231* (0.103)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	0.033 (0.073)	0.053 (0.056)	-0.095** (0.035)	-0.081* (0.034)	-0.094** (0.035)	-0.088** (0.032)	0.162 (0.143)	0.190† (0.104)
AI mentor*(\geq median)			0.256† (0.147)	0.267* (0.107)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.046 (0.058)	-0.009 (0.053)	-0.141** (0.051)	-0.122* (0.049)	-0.140** (0.051)	-0.130** (0.048)	0.050 (0.105)	0.103 (0.092)
AI mentor*(\geq median)			0.190 (0.116)	0.225* (0.104)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Covariates selected for each model are reported in Table A4. Standard errors are clustered at the individual level.

Table A4: Double-post-LASSO selected covariates

	Table A3, Panel [X]			
	[A]	[B]	[C]	[D]
Pre-treatment performance index	Y	Y	Y	Y
Pre-treatment management practices	Y	Y	Y	Y
Sector: IT Revenue and services	Y	N	N	N
Completed college	Y	Y	Y	Y
Extraversion (Big 5 personality)	Y	Y	Y	Y
Gender	N	Y	N	Y

This table reports the double-LASSO-selected covariates for Table A3. Panel C corresponds to the results reported in Figure 3. The following variables are included in our covariate selection: pre-treatment performance, management practices (sum), gender, oldest business age, sector, children (count), age, education, GPT use, Big 5 personality (agreeableness, conscientiousness, extroversion, openness, neuroticism), and county. Additionally, we always control for randomization strata blocks and time period.

Table A5: Effects on performance
(Log-like transformation)

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	-0.015	-0.004	-0.088	-0.071	-0.091	-0.099	0.055	0.049
	(0.065)	(0.060)	(0.096)	(0.092)	(0.097)	(0.091)	(0.088)	(0.075)
AI mentor*(\geq median)			0.144	0.136				
			(0.131)	(0.118)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.024	-0.014	-0.090	-0.073	-0.093	-0.102	0.038	0.022
	(0.065)	(0.061)	(0.100)	(0.095)	(0.100)	(0.094)	(0.085)	(0.072)
AI mentor*(\geq median)			0.129	0.121				
			(0.132)	(0.119)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	-0.020	-0.007	-0.096	-0.075	-0.098	-0.103	0.055	0.048
	(0.065)	(0.061)	(0.097)	(0.092)	(0.097)	(0.091)	(0.088)	(0.076)
AI mentor*(\geq median)			0.151	0.139				
			(0.132)	(0.119)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.029	-0.017	-0.099	-0.078	-0.101	-0.106	0.038	0.021
	(0.066)	(0.061)	(0.100)	(0.095)	(0.100)	(0.094)	(0.086)	(0.073)
AI mentor*(\geq median)			0.136	0.124				
			(0.132)	(0.119)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor. Outcomes are reported for an inverse-hyperbolic sine transformation. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

Table A6: Effects on performance index component: Weekly profits

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.016 (0.077)	0.023 (0.058)	-0.123** (0.042)	-0.122** (0.041)	-0.124** (0.042)	-0.123** (0.040)	0.149 (0.147)	0.157 (0.110)
AI mentor*(\geq median)			0.274† (0.153)	0.288* (0.117)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.058 (0.065)	-0.031 (0.059)	-0.180** (0.057)	-0.165** (0.055)	-0.181** (0.057)	-0.173** (0.053)	0.048 (0.116)	0.081 (0.104)
AI mentor*(\geq median)			0.228† (0.129)	0.256* (0.117)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	0.014 (0.077)	0.021 (0.058)	-0.117** (0.043)	-0.118** (0.042)	-0.117** (0.043)	-0.115** (0.041)	0.145 (0.149)	0.158 (0.111)
AI mentor*(\geq median)			0.262† (0.155)	0.280* (0.119)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.063 (0.065)	-0.035 (0.059)	-0.172** (0.059)	-0.159** (0.057)	-0.171** (0.059)	-0.164** (0.055)	0.047 (0.117)	0.082 (0.105)
AI mentor*(\geq median)			0.218† (0.131)	0.249* (0.118)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor on weekly profits. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

Table A7: Effects on performance index component: Monthly profits

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.032 (0.077)	0.047 (0.060)	-0.122** (0.044)	-0.109* (0.043)	-0.123** (0.044)	-0.112** (0.040)	0.178 (0.147)	0.201† (0.111)
AI mentor*(\geq median)			0.304* (0.153)	0.309** (0.118)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.048 (0.064)	-0.021 (0.059)	-0.166** (0.057)	-0.145** (0.055)	-0.166** (0.057)	-0.154** (0.053)	0.052 (0.115)	0.075 (0.104)
AI mentor*(\geq median)			0.217† (0.128)	0.235* (0.117)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	0.029 (0.078)	0.046 (0.061)	-0.117* (0.046)	-0.106* (0.044)	-0.116* (0.046)	-0.106* (0.041)	0.177 (0.149)	0.205† (0.112)
AI mentor*(\geq median)			0.293† (0.155)	0.305* (0.120)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.053 (0.065)	-0.025 (0.059)	-0.155** (0.059)	-0.137* (0.057)	-0.153** (0.059)	-0.143* (0.055)	0.050 (0.116)	0.076 (0.104)
AI mentor*(\geq median)			0.204 (0.130)	0.224† (0.118)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor on monthly profits. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

Table A8: Effects on performance index component: Weekly revenue

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.004 (0.070)	0.022 (0.057)	-0.073** (0.028)	-0.050† (0.029)	-0.073* (0.028)	-0.068* (0.027)	0.078 (0.136)	0.089 (0.108)
AI mentor*(≥ median)			0.152 (0.139)	0.144 (0.110)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.051 (0.060)	-0.011 (0.056)	-0.132* (0.052)	-0.099* (0.050)	-0.132* (0.052)	-0.113* (0.049)	0.022 (0.110)	0.071 (0.100)
AI mentor*(≥ median)			0.154 (0.121)	0.172 (0.111)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	0.001 (0.070)	0.024 (0.057)	-0.075** (0.028)	-0.053† (0.030)	-0.074** (0.028)	-0.069* (0.027)	0.077 (0.138)	0.106 (0.108)
AI mentor*(≥ median)			0.151 (0.141)	0.153 (0.111)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.058 (0.061)	-0.016 (0.056)	-0.136** (0.052)	-0.105* (0.050)	-0.135** (0.052)	-0.118* (0.049)	0.023 (0.111)	0.072 (0.100)
AI mentor*(≥ median)			0.157 (0.122)	0.177 (0.112)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor on weekly revenue. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

Table A9: Effects on performance index component: Monthly revenue

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.089	0.114†	-0.073†	-0.051	-0.074†	-0.062†	0.246	0.276*
	(0.089)	(0.067)	(0.039)	(0.039)	(0.039)	(0.037)	(0.173)	(0.127)
AI mentor*(\geq median)			0.322†	0.328*				
			(0.177)	(0.133)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
B) All observations, winsorized at the 95% level								
AI mentor	-0.007	0.034	-0.109†	-0.084	-0.110†	-0.089†	0.085	0.143
	(0.066)	(0.061)	(0.057)	(0.055)	(0.057)	(0.053)	(0.119)	(0.107)
AI mentor*(\geq median)			0.194	0.228†				
			(0.132)	(0.120)				
Post-treatment obs	2,514	2,514	2,514	2,514	1,251	1,251	1,263	1,263
C) Completed panel, winsorized at the 99% level								
AI mentor	0.087	0.117†	-0.071†	-0.048	-0.070†	-0.059	0.247	0.288*
	(0.090)	(0.069)	(0.040)	(0.039)	(0.040)	(0.037)	(0.175)	(0.131)
AI mentor*(\geq median)			0.317†	0.329*				
			(0.179)	(0.136)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.012	0.032	-0.103†	-0.079	-0.101†	-0.084	0.081	0.141
	(0.067)	(0.061)	(0.059)	(0.055)	(0.059)	(0.054)	(0.120)	(0.109)
AI mentor*(\geq median)			0.183	0.220†				
			(0.134)	(0.122)				
Post-treatment obs	2,488	2,488	2,488	2,488	1,244	1,244	1,244	1,244
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor on monthly revenue. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

Table A10: Probability of 0 profits or revenue by experimental condition

	p(outcome=0)		
	Full sample	Control	AI
Monthly profits			
Survey 4	0.016	0.019	0.013
Survey 5	0.008	0.006	0.009
Survey 6	0.008	0.006	0.010
Survey 7	0.008	0.006	0.010
Weekly profits			
Survey 4	0.016	0.019	0.013
Survey 5	0.008	0.006	0.009
Survey 6	0.010	0.010	0.010
Survey 7	0.011	0.010	0.013
Monthly revenue			
Survey 4	0.014	0.019	0.009
Survey 5	0.009	0.006	0.013
Survey 6	0.008	0.006	0.010
Survey 7	0.008	0.006	0.010
Weekly revenue			
Survey 4	0.014	0.019	0.009
Survey 5	0.008	0.006	0.009
Survey 6	0.008	0.006	0.010
Survey 7	0.008	0.006	0.010

This table reports the probability of reporting 0 weekly/monthly profits/revenue in each of the four post-treatment periods.

Table A11: Calculating effect sizes in percentages

	Below median	Above median
A) All observations, winsorized at the 99% level		
Profit - Month	-8.09	15.43
Profit - Week	-7.25	13.84
Revenue - Month	-8.85	16.89
Revenue - Week	-7.36	14.04
Average profit	-7.67	14.64
Average revenue	-8.10	15.46
Average (profit and revenue)	-7.89	15.05
B) Completed panel, winsorized at the 99% level		
Profit - Month	-7.77	16.71
Profit - Week	-6.97	14.98
Revenue - Month	-8.50	18.27
Revenue - Week	-7.07	15.19
Average profit	-7.37	15.84
Average revenue	-7.78	16.73
Average (profit and revenue)	-7.58	16.29

This tables presents percent estimates of our experimental treatment effects for above and below median initial performance entrepreneurs. All outcomes were winsorized at the 99% level. For rows 1-4 and 8-11 we calculate percent estimates (1) taking the treatment effect estimates on our performance index, (2) multiplying this treatment effect estimate by the standard deviation of each outcome, and (3) dividing by the mean pre-treatment value of the outcome. Rows 5-7 and 12-14 capture the average of these effect estimates across outcomes within each panel.

Table A12: Multiple inference corrected p values

	p-values (1)	BH corrected p-values (2)
A) Main performance results		
Average treatment effect	0.34	0.34
Initial low performers	0.01	0.03
Initial high performers	0.07	0.09
Difference in HTEs	0.01	0.03
B) Dimension of heterogeneity		
Pre-treatment performance	0.01	0.04
Gender	0.25	0.38
Initial ChatGPT use	0.89	0.89
C) Mechanisms		
Management practices	0.22	0.52
Time management	0.04	0.28
Technology use	0.51	0.73
Innovation	0.52	0.73
Info: Access stigma	0.83	0.96
Info: Accessibility & quality	0.08	0.29
Info: Relevance	0.96	0.96

This table presents multiple inference corrected p values following Benjamini and Hochberg (1995) for three families of tests. Within each family a correction is applied with respect to the number of tests conducted in that family. Panel A presents p values from the results presented in Figure 3. Panel B presents corrections over three pre-specified dimensions of heterogeneity. Panel C presents results correcting over our pre-specified measures designed to tap mechanisms.

Table A13: Means and Differences for Above and Below Median Performers

Variable	Below Median	Above Median	Difference	<i>p</i> -Value
Business Questions Count	3.925	4.713	0.788	0.309
Business Questions (99 Wins)	3.800	4.389	0.589	0.331
Business Questions (95 Wins)	3.529	3.978	0.450	0.329
Sample: All treated businesses. Number of firms: 317				
Business Questions Count	4.664	5.792	1.128	0.206
Business Questions (99 Wins)	4.546	5.458	0.912	0.198
Business Questions (95 Wins)	4.248	4.954	0.706	0.178
Sample: Asked a business question. Number of firms: 267				
Human Quality Rating	2.168	2.095	-0.072	0.294
Log Number of Words	2.748	2.637	-0.111	0.136
Sample: Asked a business question. Number of firms: 267				
Challenge Focus of Business Questions	0.057	-0.191	-0.248	0.0427
Sample: Asked an above median length business question. Number of messages: 695				

Table A14: Sample above-median length questions from different points of the 100th to 50th percentiles of the distribution for our *challenge-focused* measure

%tile	Questions	Challenge
100	currently facing loses in my show shop due to low demands rendering business bankrupt.please advise	2.91
100	my hotel business has faced a drastic change and is now incurring losses no matter what changes i try to implement. how can i bring back my business to thriving	2.49
99	am having a difficult time with customers since the prices have gone up and they are requesting a bargain every time	2.42
99	i run a jewelry store near ██████████. it used to have many customers but now we barely sell. what could be the problem and solutions to it?	2.41
99	hello for the past few months am experiencing low customer turnout than my neighbouring businesses	2.36
98	i am a small poultry farmer and i keep chicken in my small farm. for past few days my hens stopped laying. what could be the problem?	1.89
96	hi, i have a beauty shop behind ██████████. at first it was the best selling shop around but now it's nearly the last. what could be the problem and how do i solve it?	1.70
96	hello, i am a poultry farmer and i have been doing it for about 2 years but unfortunately my two months chicks have been dying more often. what could be the problem... if any, what could be the solution?	1.67
95	i have a cereals shop,and the type of cereals i have have been saturated in the market,how can i go about selling them without incurring losses since the prices have dropped	1.61
92	what debt collection strategies can i use to recover debts that customers have deliberately refused to pay	1.37
92	how should i convince my once customers who left due to a certain service demand which was not fulfilled	1.38
91	i have a business competitor one in specific,,he has noticed that my business is doing well and has decided too the price of the products so i cannot get customers	1.32
91	i run a bar and restaurant near ██████████. off late my workers have been running away and though it's not my fault they aren't free to say the problem. what should i do?	1.32
82	i sell household goods in ██████████. i have a problem in keeping my customers.they keep on trying other goods at my competitors how can i overcome the challenge and win my customers back?	0.98
75	nowadays i have noticed certain trend in the flow of my customers. they tend to buy more of my new introduced products as compared with the previous existing ones though i have maintained the quality. kindly advice me, what could be wrong or rather right	0.76
62	hello.my business has been doing great lately.i do my budget well allocating money correctly.but sometimes i still find some money missing when i calculate the expenditure and profits. what might be the problem?	0.29
55	in a case where i've listed the services i offer against the amount of money a client pays after the service, do i allow clients to negotiate?	0.07

Table A15: Sample above-median length questions from different points of the 49th to 1st percentiles of the distribution of our *challenge-focused* measure

%tile	Questions	Challenge
40	my hotel business in ██████ has picked and is very fruitful, but i have a problem. my customers are demanding for some services which are somehow expensive and requires some capital to introduce. how do i go about it?	-0.38
38	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.	-0.42
27	i own a small barber shop in ██████ with about 2-4 customers a day. i am trying to find ways to promote my business. i already put up posters near my shop. what are some other ways that i could gain new customers and get the word out about my business?"	-0.79
18	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?	-1.08
17	i run a salon and barber shop in one room, recently customers have started to gain interest on massage after shave, this means there is need for expansion and more resources, kindly help me to unlock this potential.	-1.12
11	i have a small hotel at ██████ where i sell tea and chapati and some other meals. how can i motivate my customers to increase my earnings?	-1.39
9	okay so i mainly deal with coffee and at times i feel like i should add more flavors like ginger and lemon to my customers but i don't know how to do it because i feel like i will interfere with other customer's preferences. so how best should i do it to ensure equality	-1.47
7	i'm thinking of investing in a new business such as farming. what ways do i need use to determine a good venture whose capital is less than let's say 50,000 shillings? and what are some of the ventures?	-1.63
7	the demand for electronics products locally has recently went up. i am looking to expand my business by opening 2 branches. the only problem is i am worried about management of the business, i don't know what could happen. what suggestions may you have?	-1.64
5	i have a grocery shop in ██████. i offer delivery services to my customers. what else can i do to increase the number of customers?	-1.82
3	i started adding milk sales to my business, i saw there was a market gap and i wanted to add more, should i run it alongside my shop or i specialise on milk since it makes more profits?	-2.04
2	i have been operating a cereals shop for five years and it is doing very well lately for the last two years. which strategies can i use to expand the business to increase capacity to serve many customers and increase the products i'm selling?	-2.44
1	hi I operate an hardware store in my town and i have thinking of how to expand the hardware store to the neighboring town what am i suppose to consider	-2.72
1	i have ksh 20000 to expand my business which it would specialize on selling food stuffs. i would like to do this expansion in a new location. what should i consider first to maximize my profit?	-2.85

Table A16: Regressions showing lower-performing entrepreneurs ask questions about more challenging and difficult business tasks. Sample restricted to above-median length business questions.

Dependent Variable: Model:	Challenge Focused			
	(1)	(2)	(3)	(4)
<i>Variables</i>				
Above Median Performer	-0.247* (0.123)	-0.215† (0.129)		
Baseline Performance			-0.359† (0.183)	-0.483* (0.215)
Log(Num of Words)	0.021 (0.105)	-0.212 (0.137)	0.049 (0.104)	-0.203 (0.137)
Level	Message	Entrepreneur	Message	Entrepreneur
Observations	695	229	695	229
R ²	0.014	0.021	0.010	0.028

Standard errors in parentheses are clustered at the level of the entrepreneur. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). The dependent variable is our standardized measure of how *challenge-focused* the entrepreneur’s questions for the AI mentor are. Columns 1 and 3 are the level of the individual messages sent. In Columns 2 and 4 we collapse the data down to the level of the entrepreneur by selecting the least challenging question the entrepreneur asks. All models control for the log number of words. The sample in this table only includes above median (14 words or more) length messages. These messages provide more context and information which tends to result in better measurement of how challenging the business situation is.

Table A17: Regressions showing lower performing entrepreneurs ask questions about more challenging and difficult business tasks. Sample includes all business questions regardless of length.

Dependent Variable: Model:	Challenge Focused			
	(1)	(2)	(3)	(4)
<i>Variables</i>				
Above Median Performer	-0.119 (0.092)	-0.190 (0.120)		
Baseline Performance			-0.328* (0.164)	-0.450* (0.208)
Log(Num of Words)	-0.132*** (0.040)	-0.299*** (0.097)	-0.123*** (0.039)	-0.294*** (0.098)
Level	Message	Entrepreneur	Message	Entrepreneur
Observations	1,392	267	1,392	267
R ²	0.013	0.043	0.017	0.049

Standard errors in parentheses are clustered at the level of the entrepreneur. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). The dependent variable is our standardized measure of how *challenge-focused* the entrepreneur’s questions for the AI mentor are. Columns 1 and 3 are the level of the individual messages sent. In Columns 2 and 4 we collapse the data down to the level of the entrepreneur by selecting the least challenging question the entrepreneur asks. All models control for the log number of words. This sample includes all business questions asked by the entrepreneurs.

Table A18: Heterogeneous Treatment Effects by Gender

	(1)	(2)
A) All observations, winsorized at the 99% level		
AI mentor	-0.004 (0.085)	0.089 (0.065)
AI mentor*(Female)	0.125 (0.162)	-0.119 (0.112)
Post-treatment obs	2,514	2,514
B) All observations, winsorized at the 95% level		
AI mentor	-0.060 (0.070)	-0.009 (0.064)
AI mentor*(Female)	0.059 (0.123)	-0.070 (0.106)
Post-treatment obs	2,514	2,514
C) Completed panel, winsorized at the 99% level		
AI mentor	-0.005 (0.085)	0.094 (0.067)
AI mentor*(Female)	0.121 (0.165)	-0.131 (0.114)
Post-treatment obs	2,488	2,488
D) Completed panel, winsorized at the 95% level		
AI mentor	-0.061 (0.070)	-0.010 (0.064)
AI mentor*(Female)	0.049 (0.126)	-0.081 (0.108)
Post-treatment obs	2,488	2,488
Time FE	Y	Y
Rand. block FE	Y	Y
Pre-treat. covariates	N	Y

This table reports heterogeneous treatment effects of the AI mentor by gender. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1 reports a base model of heterogeneous effects without covariate adjustment. Col. 2 reports covariates selected using a double-LASSO procedure. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Standard errors are clustered at the individual level.

Table A19: Heterogeneous Treatment Effects by Initial ChatGPT Use

	(1)	(2)
A) All observations, winsorized at the 99% level		
AI mentor	0.031 (0.075)	0.057 (0.063)
AI mentor*(Freq. ChatGPT use)	0.029 (0.199)	-0.022 (0.121)
Post-treatment obs	2,514	2,514
B) All observations, winsorized at the 95% level		
AI mentor	-0.036 (0.063)	-0.014 (0.058)
AI mentor*(Freq. ChatGPT use)	-0.010 (0.145)	-0.070 (0.126)
Post-treatment obs	2,514	2,514
C) Completed panel, winsorized at the 99% level		
AI mentor	0.028 (0.076)	0.058 (0.064)
AI mentor*(Freq. ChatGPT use)	0.029 (0.200)	-0.017 (0.123)
Post-treatment obs	2,488	2,488
D) Completed panel, winsorized at the 95% level		
AI mentor	-0.042 (0.064)	-0.018 (0.059)
AI mentor*(Freq. ChatGPT use)	-0.009 (0.146)	-0.071 (0.127)
Post-treatment obs	2,488	2,488
Time FE	Y	Y
Rand. block FE	Y	Y
Pre-treat. covariates	N	Y

This table reports heterogeneous treatment effects of the AI mentor by GPT use, reported as whether a participant reported using ChatGPT at least once a week prior to treatment. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1 reports a base model of heterogeneous effects without covariate adjustment. Col. 2 reports covariates selected using a double-LASSO procedure. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Standard errors are clustered at the individual level.

Table A20: Effects of AI Mentor (DID specification)

	Full sample				Below median		Above median	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A) All observations, winsorized at the 99% level								
AI mentor	0.050 (0.055)	0.050 (0.056)	-0.092** (0.033)	-0.068* (0.034)	-0.092** (0.033)	-0.092** (0.033)	0.191† (0.105)	0.188† (0.106)
AI mentor*(\geq median)			0.283* (0.110)	0.236* (0.111)				
N Entrepreneurs	637	637	637	637	318	318	319	319
B) All observations, winsorized at the 95% level								
AI mentor	0.005 (0.053)	0.005 (0.054)	-0.132** (0.048)	-0.100* (0.050)	-0.132** (0.048)	-0.132** (0.048)	0.143 (0.094)	0.141 (0.095)
AI mentor*(\geq median)			0.275** (0.106)	0.212* (0.106)				
N Entrepreneurs	637	637	637	637	318	318	319	319
C) Completed panel, winsorized at the 99% level								
AI mentor	0.049 (0.057)	0.049 (0.057)	-0.089** (0.033)	-0.074* (0.034)	-0.089** (0.033)	-0.089** (0.034)	0.186† (0.108)	0.186† (0.108)
AI mentor*(\geq median)			0.274* (0.113)	0.246* (0.113)				
N Entrepreneurs	622	622	622	622	311	311	311	311
D) Completed panel, winsorized at the 95% level								
AI mentor	-0.001 (0.054)	-0.001 (0.054)	-0.130** (0.049)	-0.108* (0.050)	-0.130** (0.049)	-0.130** (0.049)	0.126 (0.096)	0.126 (0.096)
AI mentor*(\geq median)			0.256* (0.108)	0.216* (0.108)				
N Entrepreneurs	622	622	622	622	311	311	311	311
Time FE	Y	Y	Y	Y	Y	Y	Y	Y
Rand. block FE	Y	Y	Y	Y	Y	Y	Y	Y
Pre-treat. covariates	N	Y	N	Y	N	Y	N	Y

This table reports the performance effects of the AI mentor from a difference-in-difference specification. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. All tests are two-tailed, and no adjustments have been made for multiple comparisons (see subsection E.5 for a details). Cols. 1-2 report average treatment effects. Cols. 3-4 report heterogeneous treatment effects based on whether the entrepreneur has pre-treatment performance that was above median. Cols 5-6 report average effects on below-median entrepreneurs, and 7-8 report effects on above-median entrepreneurs. Panels A and B report effects using all post-treatment observations, and panels C and D report effects for participants who complete the entire panel. Panels A and C report effects winsorized at the 99% level, and panels B and D report effects winsorized at the 95% level. Covariates (in even-numbered columns) were selected using a double-LASSO procedure. Standard errors are clustered at the individual level.

C Experimental Design and Timeline

Our field experiment lasted approximately five months, and involved 640 Kenyan entrepreneurs and seven separate performance surveys. It was structured in three stages:

1. **Recruitment and pre-treatment Surveys:** Initially, entrepreneurs were recruited online and screened using three pre-treatment surveys.
2. **Assignment to Interventions:** Participants were then assigned either to an AI mentorship program or to receive a standard business training guide.
3. **Performance Tracking:** Finally, we assessed each participant’s subsequent business performance through four post-treatment surveys.

We provide details on these stages below.

Stage 1: Recruitment and pre-treatment Surveys

Participant recruitment took place over the Meta ad platform (Facebook penetration in Kenya is over 80% (Wamuyu, 2020)). A partner survey organization created advertisements offering a paid survey targeting entrepreneurs and businesses. The advertisements were shown throughout Kenya and the only restriction on the advertisement was it was only to be shown to individuals over the age of 18 (see Figure A1).

Individuals who clicked on our ad were directed to a Qualtrics survey, which was the first of three pre-treatment surveys we conducted with participants. These pre-treatment surveys aimed to screen out individuals: (i) who did not own and operate enterprises; (ii) who would be likely to attrit over the long time horizon of our study; (iii) who were inattentive respondents.

Our analytic sample predominantly consists of relatively young businesses, with 61% having been in operation for one year or less, and only 16% having been operating for over three years. The businesses were primarily concentrated in three sectors: Food and Beverage (27%), Clothing (20.8%), and Agriculture (18.9%). A majority of these businesses (64%) operated without any employees. Pre-treatment the median firm in our sample earned a monthly profit of 14,000 Kenyan Shillings (about \$90 USD) and had a monthly revenue of 30,000 Kenyan Shillings (about \$190 USD).

Participants came from 44 of Kenya’s 47 counties, with the largest number of participants coming from Nairobi (133 participants), Kiambu (53 participants), and Nakuru (47 participants). Nairobi’s 4-million-person capitol has the highest concentration of study participants (20%). Participants in our sample generally exhibited a high level of education, with only 10% not having any college education. The sample demonstrated a low level of gender diversity, with women accounting for only 32% of the participants, although this may in part reflect a lower level of business ownership among women in Kenya (The World Bank Group, 2018). Additionally, the average age of participants in our study was 26 years old. See Table A1 for additional details on our sample.

Participants who completed the first survey were contacted via SMS messages through a “Shortcode” (a shortened phone number used to bulk send text messages) and also over email inviting them to participate in a second survey. These individuals were sent invitations to a second survey via a personalized Qualtrics link that could only be used one time.²¹ Those completing the second survey were sent a third pre-treatment survey over WhatsApp. Following the completion of the pre-treatment surveys but prior to random assignment, we pre-registered the following exclusion criteria:

1. Participants who do not run a business or do not state that they are willing to take subsequent surveys
2. Participants who tried to take any Qualtrics survey more than once, as identified by Qualtrics duplicate score or by duplicated phone numbers or emails
3. Participants who fail the following attention check:

The next question is about the following problem. In questionnaires like ours, sometimes there are participants who do not carefully read the questions and just quickly click through the survey. To show that you read our questions carefully, please enter orange as your answer to the next question. What is your favorite color? {red; yellow; green; blue; black; orange; white; purple; pink}

4. Participants who stated that their weekly profits (revenue) were greater than their monthly profits (revenue), or who stated that their profits were greater than revenue (separately for weekly and monthly measures).

²¹All subsequent survey waves also used similar one-time-use personalized Qualtrics links.

Stage 2: Treatment

After the third pre-treatment survey wave, we randomly assigned participants to one of the following experimental conditions.

1. **AI-mentor.** Participants assigned to the AI-mentorship condition received access to our WhatsApp-based AI-mentor. The AI mentor was powered by GPT-4 and designed to give business mentorship and advice to Kenyan small and medium businesses.
2. **Control.** Participants assigned to the control condition received access to widely used business training materials developed for small and medium businesses by the International Labour Organization.

Randomization was stratified on Gender \times (Quartile of pre-treatment business performance), for a total of $2 \times 4 = 8$ strata.

Onboarding survey. Before participants gained access to the business guide or the AI mentor, they were required to complete a brief *Onboarding Survey*. The first page of the onboarding survey provided a concise summary of the expected time to complete the survey and the mobile airtime compensation they would receive upon completion, which offset the data costs of taking the online survey. Notably, the first page did not disclose the participant's treatment group. Participants who did not proceed beyond the first page, where treatment status was undisclosed, are considered pre-treatment dropouts. Our analytic sample is comprised of 640 participants who advanced to the second page of the Training Survey, where they were informed of their treatment status. These participants had already successfully completed three rounds of pre-treatment surveys and passed pre-registered attention checks.

Stage 3: Post-treatment surveys

Following treatment, we conducted four additional survey waves in which participants were contacted via SMS messages. Each of these waves contained a battery of questions exploring mechanisms in addition to our set of performance measures. See Appendix E for an overview of the contents of each survey.

Survey Incentives

Participants completing the first survey were sent a small airtime transfer of KSH 30 as a participation incentive. These credits could be used to purchase phone credits and cellular data to access the internet. To reduce attrition we used an increasing survey incentive scheme: The second pre-treatment survey paid KSH 50, and the third paid KSH 100. All post-treatment surveys paid KSH 200.

Attrition

Our analytic sample, defined by pre-registered exclusion criteria (see Appendix C) is comprised of 640 participants. A total of 18 (3%) of participants attrit from the survey post-treatment (i.e., did not complete our final survey, but may have completed at least one pre-treatment survey).²² The final row of Table A1 reports the final sample size by experimental condition. In total 11 participants attrited from the “AI” condition and 7 attrited from the control condition. The difference in attrition across conditions is 1.2 pp ($p = 0.35$).

²²Studies such as Banerjee et al. (2015) and Bossuroy et al. (2022) report attrition rates of approximately 9%.

D Detailed Intervention Overview

AI mentor

Our AI mentor was powered by the GPT-4 and Whatsapp APIs. WhatsApp has very high adoption in Kenya (e.g., Wamuyu (2020) report that 89% of Kenyans use WhatsApp and 81.7% use Facebook). Participants interacted with the AI mentor over WhatsApp by texting our WhatsApp number. While the number was public (anyone can text the number), we programmed the AI mentor only to respond to participants on a designated access list. Thus, there was no risk that control participants could interact with the AI mentor. The AI mentor used the standard GPT-4 corpus and was contextually tailored using prompt engineering alone. All parameter values (e.g., “temperature”) were set to GPT-4’s default values.

Development of the AI Mentor

We used an iterative process to develop our AI mentor over the course of several months. This process was structured into three, interconnected phases.

1. In the first phase, we worked with a team of research assistants to craft prompts aimed at providing strategic advice to Kenyan entrepreneurs. This involved a two-step process: first, a team member posed basic business questions to GPT-4. Then, another evaluated the responses for quality and relevance. We cycled through this process repeatedly, refining the prompts in each cycle based on the feedback received in the cycle prior.
2. The second stage involved a team member manually forwarding questions asked by a pilot sample of Kenyan entrepreneurs to GPT-4. As part of this process, our team and the sample of Kenyan entrepreneurs evaluated the quality and form of the GPT-produced responses. As above, we progressively revised our system prompt based on feedback.
3. The third and final cycle took place after we had built an early version of the AI mentor that was able to communicate with participants over WhatsApp. Here, we recruited a new sample of Kenyan pilot participants who were given early access to the AI

mentor. In addition to asking questions directly to AI mentor, these pilot participants completed surveys that captured more detailed feedback on the tool, which we again used to iteratively revise our system prompt.

Based on this process, our final product was an AI mentor instructed to adhere to the following guidelines:

- For each query, the AI mentor provided 3 to 5 tailored, actionable, enumerated strategies.
- After each response, the AI mentor prompted further engagement with: “Enter a number to explore more about one of these points on [topic name], or ask a new question,” ensuring conversational clarity and flow.
- When a user replied with a number, indicating interest in a specific strategy, the AI mentor offered 3 to 5 additional actionable tips related to that strategy.

The AI mentor was endowed with the personality of “an expert business mentor specialized in guiding Kenyan entrepreneurs with limited resources” whose goal was to “improve business performance and profitability by providing actionable, contextually relevant, specific, detailed advice that produces sustained improvements in competitive advantage.”

AI Mentor System Prompt

Below is the final version of the system prompt that we provide to GPT-4 in order to create the AI business mentor:

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.

ILO guide

In the control group participants were provided with a guide developed by the International Labor Organization. Specifically, we texted (over WhatsApp) the first three PDFs (out of a series of 9 PDFs in the guide) to participants. We also provided a link where participants could access the other PDFs. The full series of ILO guide PDFs can be accessed at https://www.ilo.org/empent/areas/start-and-improve-your-business/WCMS_192062/lang--en/index.htm.

Training Materials

AI Mentor Training

Here at O-Survey, we have built a special robot that can give you advice to help you with your business. This AI-powered robot, or “AI mentor chatbot” is designed to act like a business mentor that you can chat with any time you want. It provides custom advice to improve the performance of your business. We are going to give you access to the AI mentor chatbot for free over WhatsApp. You can interact with the AI mentor chatbot on the same WhatsApp number that sent you this survey. You may find it helpful to chat with this AI mentor at least a few times a week (you can talk to it more if you want). We will send you a message over WhatsApp in 1-2 days when you can begin talking with the AI mentor. Now you will spend 5-10 minutes learning about how to make the best use of the AI mentor chatbot.

Here are a few things you should know about the AI mentor chatbot:

- The advice is provided by an AI system (not by a human mentor). Sometimes, the advice may be incorrect or bad. Make sure to think about the advice before you use it.
- The AI mentor chatbot doesn't know what's happening in the world right now. So, you can't ask it things like “How hot was it yesterday?”
- Sometimes, it might take up to 2 minutes for the AI mentor chatbot to reply. Just wait a bit before asking another question.

Here are some tips to get the best advice from our AI mentor chatbot:

1) Be detailed: The more you tell the AI mentor chatbot about your situation, the better it can help. For example, 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?”

Here is another example of a good question: “I own a small hair salon in Nairobi with about 2-4 customers a day. I am trying to find ways to promote my business. I already put up posters near my shop. What are some other ways that I could gain new customers and get

the word out about my business?”

This question is good because it provides details on who is asking the question (“I own a small hair salon in Nairobi with about 2-4 clients a day”), and it provides details on the problem and what you have already tried (“I already put up posters near my shop”). Then it ends with the question (“What are some other ways that I could gain new customers and get the word out about my business?”).

2) Ask follow-up questions: When the AI mentor chatbot gives you advice, you can ask follow-up questions. Here is an example response from the AI mentor chatbot to the previous question about promoting a business:

Question: I own a small hair salon in Nairobi with about 2-4 clients a day. I am trying to find ways to promote my business. I already put up posters near my shop. What are some other ways that I could gain new customers and get the word out about my business?

Answer: “Here are three practical strategies:

- Leverage social media: Create accounts on popular platforms like Facebook or Instagram. Regularly post attractive pictures of hairstyles, promotions, or events at your salon to engage with potential clients.
- Offer referral incentives: Encourage existing customers to refer friends by giving them discounts or other rewards for each successful referral.
- Host community events: Organize small gatherings like beauty workshops or charity events at your salon so people can experience the quality of service firsthand while contributing positively to the local community.

Please let me know if you’d like more information about any of these marketing techniques (1-3) or if you have another question related to growing your business.”

You will get the most benefit if you ask follow-up questions to the AI mentor chatbot. For example, you could say “2. What are some ways that I can use referral incentives to get new customers to my barbershop?” to learn more about the second option.

To summarize, to get the best help from our AI mentor chatbot:

- Remember to give lots of details about your situation
- Ask follow-up questions about the advice it gives.

In addition to asking followup questions, you can also ask for a new question at any time.

Remember, you will be given access to the AI Mentor over WhatsApp in 1-2 days.

Control (ILO Guide) Training

We would like to help you access resources that could help your business. Specifically, we would like to provide you with business training manuals developed by the International Labor Organization (ILO). Note that we are not affiliated with the ILO.

We request that you please do not share the materials. To get the most out of these materials, you may find it helpful to review them at least a few times a week.

Now let's learn about the materials. Here is an overview of some of the materials we are providing:

1. A guide to coming up with new business ideas
2. A guide to starting a new business
3. A guide for developing a new business plan

While some of these materials are focused on new businesses, they can also be used to evaluate and develop current businesses as well. These materials will be sent to you over WhatsApp 1-2 days after completing the survey.

Now you will spend 5-10 minutes learning about how to make the best use of the materials.

1) Scan the materials: Each of the documents we will share has a "Contents" page that provides an overview of the contents offered in the specific guide. For example, the guide "Generate Your Business Idea" has the following sections:

- PART I – YOU AS AN ENTREPRENEUR
- PART II – A MARKET IS WAITING FOR YOU
- PART III – YOUR OWN BUSINESS IDEA LIST
- PART IV – THE BEST BUSINESS IDEAS FOR YOU
- PART V – YOUR OWN BUSINESS IDEA

Other guides allow you to evaluate your business ideas and more!

2) Business planning: Creating a business plan can help you manage your business, understand who your customers are, what your products or services are, and much more! Careful planning will help you get the most out of your business. The business plan includes guides on:

- Market Research
- Sales Estimation
- Product Costing
- Sales Planning
- Start-up capital

The more details of a plan you develop, the better you will be able to understand the challenges your business will face.

After reading the “Start Your Business” guide you will know how to:

- Describe the content of a Business Plan
- Consolidate your business idea
- Translate your business idea into a completed Business Plan
- Assess your readiness for starting a business

And much more! Remember, you will be sent the materials over WhatsApp in 1-2 days.

Reminders

In both experimental conditions, we provided participants with intermittent reminders to engage with the business manual or AI mentor for each condition.

E Pre-analysis Plan

E.1 Overview of Pre-registration

We pre-registered our study design and analysis on osf.io after the third pre-treatment survey. The following dimensions were pre-registered:

Outcomes: subsection E.3 outlines all of our pre-registered survey outcomes:

- (i) Performance (revenue and profits)
- (ii) Management Practices
- (iii) Time Management
- (iv) Technology use
- (v) Innovation activities
- (vi) Information: Access Stigma
- (vii) Information: Relevance
- (viii) Information: Accessibility and Quality

We also pre-registered the construction of our performance index and the level of winsorization we would apply our outcome variables. A more detailed overview of our survey measures is included below.

Statistical analysis: We pre-registered our primary ANCOVA regression specification (see below).

Heterogeneous effects: We pre-registered three dimensions with respect to which we would measure treatment effect heterogeneity:

- (i) Initial performance
- (ii) Gender
- (iii) Pre-treatment ChatGPT use

We elaborate on each component of the pre-registration below.

E.2 Econometric Strategy

Our main pre-registered estimation strategy is an ANCOVA variant of a simple OLS regression (McKenzie, 2012):

$$y_{i,t} = \alpha + \beta \text{AI}_i + \gamma y_{i,-t} + \zeta X_{i,-t} + \sum_S \theta_s \mathbb{1}(i \in s) + \sum_T \tau_j \mathbb{1}(j = t) + \varepsilon_{i,t} \quad (1)$$

where $y_{i,t}$ is an outcome variable measured for participant i at time t . α is the constant, and $\text{AI}_i \in \{0, 1\}$ is a dichotomous dummy taking a value of 1 if i is assigned to receive access to the AI mentor. β captures the average causal effect of access to the AI mentor in post-treatment periods. $y_{i,-t}$ is the pre-treatment values of outcome y in pre-treatment periods ($-t$). $X_{i,-t}$ is a vector of baseline control variables. θ_s captures stratification dummy variables and τ_j captures post-treatment time dummy variables. $\varepsilon_{i,t}$ is the error term. Standard errors are clustered at the individual level.

Heterogeneous treatment effects: We estimate heterogeneous treatment effects for three pre-registered dimensions: gender, baseline business performance, and whether participants report frequent use of ChatGPT at baseline. In practice, we operationalized baseline business performance as a median split of a standardized performance index over the three pre-treatment periods, and we operationalized frequent ChatGPT use as whether the person uses ChatGPT at least once a week. We did not provide the heterogeneous treatment effects estimating equation in our pre-analysis plan but for clarity we report the parallel heterogeneous treatment effects estimation equation to Equation 1 below:

$$y_{i,t} = \alpha + \beta_1 \text{AI}_i + \beta_2 \text{Het}_i + \beta_3 (\text{AI}_i \times \text{Het}_i) + \gamma y_{i,-t} + \zeta X_{i,-t} + \sum_S \theta_s \mathbb{1}(i \in s) + \sum_T \tau_j \mathbb{1}(j = t) + \varepsilon_{i,t} \quad (2)$$

where Het_i is the relevant dimension of heterogeneity, and β_3 captures the heterogeneous treatment effect.

Covariate adjustment: We follow a principled approach to covariate selection using the

double post-LASSO method (Belloni, Chernozhukov, and Hansen, 2014). This approach involves first running a LASSO to identify outcome-relevant covariates, and second, to identify variables predictive of treatment assignment. The union of this set of covariates is then included as controls in our regression analysis. The same double-LASSO-selected covariates are used in the average treatment effect and the heterogeneous treatment effect variations of each model. Additionally, we always control for randomization strata blocks and time period.

The following pre-treatment variables are included in our covariate selection: performance, management practices (sum), oldest business age, sector, children (count), age, education, GPT use, “Big Five” personality measures (agreeableness, conscientiousness, extroversion, openness, neuroticism), and county. See Table A4 for details covariate selection for our main regression results.

E.3 Overview of Measures

Business performance. We measure business performance using a combination of measures of revenue and profits. Following Anderson and McKenzie (2022), we apply the following transformations:

1. Winsorize the outcome at the 95th percentile (and 99th percentile in a separate specification).
2. After applying the step above, generate an average z -score following Kling, Liebman, and Katz (2007) by subtracting the mean and then dividing by the standard deviation of the winsorized vector of pre-treatment outcomes. Finally we average across the three performance measures: (a) Weekly sales; (b) Weekly profits; (c) Monthly sales; (d) Monthly profits.

Pre-treatment performance. We calculate pre-treatment performance by calculating the average value of the four performance outcomes (weekly and monthly profits and revenue) across the three pre-treatment periods. We then calculate a z -score of winsorized average performance for each of the four measures. Finally, we calculate the average of these four standardized pre-treatment performance measures. For log-like pre-treatment performance measures the first transformation applied is the inverse hyperbolic sine transformation.

Business practices. We create an average z score out of a set of sixteen management practices based on McKenzie and Woodruff (2017) and 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, mentorship, 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 judgement

- [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 judgement 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}

E.4 Explanation of Preferred Functional Forms

In our pre-registration, we pre-specified four different variations on a performance index: 95 and 99% winsorization and a log-like inverse hyperbolic sine transformation $y' = \log(y + (y^2 + 1)^{1/2})$ transformation of these two winsorizations. In the paper body, we focus on the raw outcome variable and report the raw unit results as our preferred specification for two reasons.

First, recent work has shown that the inverse hyperbolic sine transformation (and other similar log-like transformations) is highly sensitive to artefactual statistical features like scaling of the outcome (e.g., measuring performance in units of thousands of KSH), and also the presence of zeros (Chen and Roth, 2023; McKenzie, 2023). We observe exactly this sensitivity to zeros in our data. Figure A12 shows the distribution of our performance indices, winsorized at the 95% level on the left and the 99% on the right. The top row depicts raw units and the bottom depicts log-like units. We observe that for the log-like index, there is an extreme left tail that arises from zeros in the performance distribution, with outcome values close to or above -10 standard deviations.

Second, for the purpose of studying inequality, 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.5 Hypothesis Testing and Multiple Comparisons

We address elevated false discovery rates resulting from multiple comparisons through the following approaches.

First, our analysis condenses our four distinct performance measures—weekly profits, monthly profits, weekly revenue, and monthly revenue—into a single pre-registered performance index. This approach not only reduces noise and enhances our statistical power but also decreases the number of hypotheses tested in the paper, thereby lowering the false discovery rate across what would otherwise be multiple tests for different performance measures (Anderson, 2008; Kling, Liebman, and Katz, 2007).

Second, we apply multiple inference corrections following Benjamini and Hochberg (1995) across three key test families: (1) our performance ATE results presented in Figure 3; (2) our three pre-registered dimensions of heterogeneity (pre-treatment performance, pre-treatment ChatGPT use, and gender); (3) and our seven indices of mechanism questions. Table A12 reports the original and BH-corrected p -values. Our main findings—such as the difference between high and low performer HTEs and the negative ATE for low performers—are robust to corrections for an elevated false discovery rate due to multiple comparisons.

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 mentor 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 mentor, 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 mentor, 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 Accounting for 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 mentor, 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 mentor’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 mentor. To address this, we developed our AI mentor in such a way that it only responds to messages from individuals on a pre-specified access list. Only members of the research team were able to modify this access list.

Performance Spillovers

Another concern relates to performance spillovers within our experimental sample. For instance, the heterogeneous treatment effects we observe could potentially be explained by the high-performing entrepreneurs in our sample capturing profits/revenue from the low-performing entrepreneurs in our sample. However, we consider this unlikely for two reasons. First, our advertisement on Meta targeted a broad audience, drawing participants from across Kenya, a country of over 50 million people where Facebook has very high market penetration (Wamuyu, 2020). Consequently, our recruitment strategy yielded a geographically diverse sample. Table A1 illustrates the distribution of respondents over Kenyan counties (looking only at those counties with over 50 respondents). Second, negative spillovers are overwhelmingly more likely to occur between individuals who are located in the same geography *and who work in the same industry*. In Figure A3, we further stratify our sample by county and sector. Because our sample is drawn from all over Kenya, every sector-location combination has fewer than 40 participants (note that some participants have businesses in multiple sectors). Only two sector-location combinations have over 20 individuals: Nairobi × Food and Nairobi × Clothing. As a point of reference, Nairobi has a population of over

4,000,000 and the Nairobi metropolitan area spans over 250 square miles.

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 mentor 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 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 mentor 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 mentor 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 [REDACTED]. 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 [REDACTED], 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, [REDACTED]. 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 [REDACTED], 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 [REDACTED], 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, [REDACTED]. 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 mentor 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 mentor 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 mentor 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 mentor also discussed the risks associated with implementing new business strategies, with the AI mentor 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.

I Text Classification and Coding

This section describes the coding practices used to classify and rate interactions with the AI mentor. Two human RAs manually classified the thousands of messages sent by entrepreneurs to our AI mentor into one of three classes: {*Business messages*, *Not Business messages*, and *Other Questions/Not a Question*}. The human coders further assigned a quality rating score to each business question. Our text analyses focus on the 1,392 messages classified as business questions.

Question Classification Protocol

The following definitions were used to manually code each message from an entrepreneur into one of three classes: Business Questions, Homework Questions, and Other Questions/Not a Question. Examples of messages that were assigned in each label can be found below.

Text	Label
hi AI Mentor. I've had a problem in my grocery business as money transfer companies have increased the transactions costs. usually my customers do not prefer cash transaction but <i>M-pesa</i> transactions. how can I deal with this please?	Business
write me a business proposal of a grocery store	Business
Discuss in detail the following privacy issues: (i) The key roles and responsibilities in risk management (ii) The Key Risk Indicators (KRI) in risk management (iii) The various risk management frameworks...	Homework
Vulnerability is defined as a known weakness or fault in a system that exposes information to attack or damage.Using the list below describe ten vulnerabilities associated with each (i) Hardware (5 Marks) (ii) Software (5 Marks)...	Homework
what is the relationship between the Australian and British flag?	Non-business
since I am a kenyan citizen what is it required of me to secure a place in one of the above mentioned universities	Non-business

Business Questions

Human coder's subjective judgments were used to classify messages as "business" or "not business. When evaluating messages, coders read through each conversation in order so that contextual information from the conversation could be used to improve the quality of the

classification scheme. Broadly, messages were classified as “business” if they were related to an entrepreneur’s current or prospective business activities, or brainstorming ideas for a new business. General guidance for this classification includes the following:

1. The intent behind asking the question can be traced back to business needs.

Example Questions:

- a) *I own a business in ██████ selling ladies bags. The sales are slow. How do I increase my sales?*
- b) *I have chicken they produce alot of eggs, hatching is not a problem but when the chicks finishes one week approximately they start to die in masses. What can I do to make them have a healthy life?*

2. The user mentioned an existing business or intent to start/ideation about another business.

Example Questions:

- a) *Hi, what is the business idea that I can implement in an environment full of university students?*
- b) *Some new business ideas someone could think of*

Homework Questions

Homework questions are those questions that human coders consider (given the context of surrounding questions) to be answering an exam or assignment. Some of these questions have key identifying information that is characteristic of homework questions like the presence of points awarded.

Other Questions or Not a Question

These include questions on non-business related issues and text that is not a question. For example, questions focused on personal life or general questions users ask about politics. It also includes text that appears to be accidentally sent to the AI mentor but is clearly not a question.

Resolution Process for Question Category Label Discrepancies

The raters reconciled any discrepancies by re-classifying messages that received conflicting classifications. We provide examples of this reconciliation process below. In the first example, rater 1 classified “What is petty cash book?” as not a business question, while rater 2 classified it as a business question. When a discrepancy occurred, the raters re-classified questions collaboratively using the following ad-hoc approach: First, each rater gives the motivations behind their original classification. Then, to obtain a more comprehensive understanding of the question in dispute, the raters re-examined both the sequence of messages that preceded and those that superseded the focal message, as well as the corresponding AI mentor responses. With a clearer understanding of the surrounding messages and the circumstances, the resolution process ends with a consensus-based final message classification.

Question	Rater 1	Rater 2	Resolution
what is a petty cash book?	Non-Business	Business	Business
I had a customer today who paid half the money but did not say when they were going to finish the pay. How do I get a date or assurance of payment?	Business	Non-Business	Business
how often should you review and update your tone of voice to align with your evolving brand identity?	Business	Homework	Business
what are advantages of entrecomp in entrepreneurship	Non-Business	Business	Business
what is the best type of sawdust to use. is it from hardwood trees or softwood trees?	Business	Other	Business

Quality Rating Protocol

Business messages were also annotated with a quality rating of {Low, Medium, or High}. As with the category rating described above, quality ratings were made holistically based on the quality of the question situated within the broader context of the conversation between the AI mentor and the participant. As a result, very short questions were not necessarily rated as low quality, and very long questions were not necessarily rated as medium/high quality. In the case of quality label discrepancies, the resolution process described above in detail was followed.

In what follows we provide examples of messages that were given different quality ratings, and outline common characteristics of Low, Medium, and High-quality questions. Note that these are heuristic guidelines, as opposed to definitive rules that were used while making the classifications.

Text	Rating
how should I set price that can fit for all clients?	Low
maximize profit	Low
how to increase customer relations	Low
how can I improve my sales per day	Low
I want to get more yield from my small poultry farm of 10 chicken. what should I do?	Medium
I have been selling honey but the profits are going low due to more people doing the same business in my area. What else can I sell together with honey to get more profits and attract more customers?	Medium
what is the best business to start with a capital of 10,000?	Medium
which species of maize can be grown in low rainfall areas?	Medium
I own a small Tailoring shop in ██████████ with a few clients in a day. I am looking on ways to promote my business. I have talked to a few friends about the business. What are some ways that can help me grow my business??	High
I own a hardware shop in ██████████. I serve customers with wide range of building materials. I have done delivery services in an attempt to attract more customers. What are the other ways that I can do to attract more customers into my hardware shop?	High
I have capsicum bell pepper which I am planning to plant on my 50×20 very soon as it matures from my seedbed. How do I plant capsicum bell pepper and how do I make high yields?	High

Low Quality

Low-quality questions generally had the following features:

- The message is not coherent and logical.
- No information about the business context and setting is provided.
- The AI Mentor will be unlikely to understand and effectively answer the question.

Medium Quality

Relative to a low-quality message, medium-quality messages tended to have the following characteristics:

- The question is slightly more coherent and logical.
- Some information about the business/setting is provided. Contextual information about the problem the entrepreneur is facing may be provided, but with limited detail.
- The AI Mentor will be more likely to understand and effectively answer the question.

High Quality

Relative to low and medium-quality questions, high-quality questions had following characteristics.

- The message is comprehensive, i.e., it is self-sufficient and does not require additional information for contextual understanding.
- Detailed information about the business and setting is provided.
- When relevant, messages may provide information on what the entrepreneur has already tried.
- The AI Mentor is very likely to understand and effectively answer the question.

Note again that high-quality questions may be short and well-defined conceptual questions; they are not necessarily longer than low or medium-quality questions.

J Word Embedding Analysis of Question Challenge Focus

To generate our “challenge-focused” measure, we first generated a 1,536-dimensional word embedding vector for each question sent to the AI mentor using OpenAI’s ada-002 model. While these vectors tell us where each question sits in semantic space, they do not reveal if the question is focused on fundamental business challenges or not. To measure the extent to which any particular question is focused on difficult-to-fix business tasks, we first calculate the word embedding vectors corresponding to two benchmark questions covering similar topics, one of which is written from the perspective of a business facing difficult questions and one of which is written from the perspective of a business that is experiencing success. The benchmark question for struggling businesses is:

My business is really struggling and facing major challenges. There is lots of competition and its hard to raise capital. Everything is really tough and there are lots of constraints. I need help trying to fix these major problems from financing constraints to having no time to no one wanting to buy my products in order to turn my business around and hopefully find some success. I am not getting paid, my profits are less than my expenses, and I feel stuck. I have basically no customers and have a hard time finding more clients.

whereas the benchmark question for successful businesses is:

My business is doing well but I want to start growing and improving even more. I want to expand and raise more capital, it could be debt equity or microfinance, to make my business thrive. It is hard work, but I think there is a lot of possibility for my business. I need help trying to hire more people, improve my prices, costs and budgeting, and market my business and get some financing to become an even bigger success. I am making money but think I have the potential to make even more. I have some customers but think I can get even more.

We proceed to project the vector representing each business question in our sample onto the line connecting the pair of vectors representing each benchmark question. Since both

benchmark questions cover similar topics, their corresponding vectors should occupy similar positions in semantic space, making it possible for us to isolate the extent to which each question in our data is about challenging tasks and circumstances. Table A14 and Table A15 show examples of questions that were rated as above- and below-median on our *challenge-focused* measure, respectively. It is apparent that this measure is not merely capturing differences in, e.g., how finance or customer focused a question is; there are examples on both ends of the spectrum covering each topic. This fact suggests that our *challenge-focused* measure is indeed capturing differences in whether a question is focusing on fundamentally challenging and difficult business tasks.

Using this *challenge-focused* measure, we are able to run regressions that test whether low performers are more likely to ask challenge-focused questions. In Table A16 we report four regressions that are estimated on the subset of questions that are above-median-length.²³ In all models the dependent variable is our *challenge-focused* variable. In Model 1 we run the regression at the message-level and cluster standard errors at the entrepreneur-level. We find high performers (above median) asked questions that were -0.247 standard deviations less challenge-focused; this relationship is statistically significant at the 5% level. Model 2 collapses the data down to the entrepreneur-level by focusing on the least challenging question each entrepreneur asked.²⁴ Model 2 yields similar results to Model 1; high performers asked questions that were -0.215 standard deviations less challenging; this thought the effect is only statistically significant at the 10% level. Models 3 and 4 use a continuous measure of baseline performance and yield similar results and significance levels. All models control for the logged number of words in the question.

In Table A17, we show that the findings in Table A16 hold when we use questions of all lengths, as opposed to only questions that are above-median-length. While in Models 1 and 2, the relationship between binarized baseline performance and our *challenge-focused* measure is

²³Many questions in our data are quite short, often because they are following up an earlier, longer question. Because of this, we initially focus our analysis on questions of length greater than or equal to 14 words, the median question length in our data. These longer questions provide enough semantic context to generate reliable embedding vectors.

²⁴Given that entrepreneurs can ask multiple questions, it is possible that they ask an especially hard question, realize it is beyond either their or the AI mentor’s capabilities, and then turn to an easier question where the AI mentor can provide useful advice. This suggests focusing on each entrepreneur’s least challenging question is appropriate.

no longer significant, the sign is still negative. Furthermore, the relationship between baseline performance and our *challenge-focused* measure becomes significant at the 5% level again in Models 3 and 4 when we use the continuous, rather than binarized measure of baseline performance. Unlike the models reported in Table A16, in these models the coefficient on the logged number of words is negative and extremely significant. In other words, shorter questions are more likely to be marked as more challenging. While potentially concerning, the fact that the relationship between baseline performance and question difficulty is consistent across the models reported in Table A16 and Table A17 suggests that our results are not simply an artifact of question length.