The Workforce Selection Effects of Training with AI: Evidence from a Field Experiment

Emaad Manzoor,[1*] Hongye Sun,[2] Kun Peng,[3] Vrinda Kadiyali[1]

[1] SC Johnson College of Business, Cornell University
[2] David Eccles School of Business, University of Utah
[3] Business School, Nanjing University

November 11, 2026

Abstract

We study the workforce selection effects of replacing human practice partners with AI in sales training. In a field experiment at a medical equipment firm, we randomize new hires to practice with AI or practice with peers. Two weeks after training completion, practicing with AI produces a larger retained workforce with lower mean extraversion and increases cumulative sales calls per recruit relative to practicing with peers. A statistical decomposition attributes 84–93% of the increase in calls to additional days of employment among introverts, partly offset by fewer days of employment among extraverts. Our findings identify workforce selection on personality traits as an important consequence of training technology.

1 Introduction

Artificial intelligence is transforming salesforce training, a $15 billion annual expenditure [Atefi et al., 2018]. Firms are increasingly replacing or augmenting human sales coaches and practice partners with AI that trains salespeople through interactions with simulated prospects [Council, 2019; Microsoft, 2025; Armstrong, 2026; LinkedIn Learning, 2025; Gonzalez et al., 2026].

Existing research on training salespeople with AI has primarily examined its effects on worker performance through human capital formation [Luo et al., 2021; Habel et al., 2026]. However, a longstanding literature in labor economics recognizes that the provision of training can also affect workforce selection, particularly through differences in worker ability [Parsons, 1972; Jovanovic, 1979; Barron et al., 1987; Acemoglu and Pischke, 1998; Autor, 2001]. Motivated by this literature,

*Emaad Manzoor is an Assistant Professor of Marketing at the SC Johnson College of Business, Cornell University. Hongye Sun is a PhD student at the David Eccles School of Business, University of Utah. Kun Peng is a PhD student at the Business School, Nanjing University. Vrinda Kadiyali is the Nicholas H. Noyes Professor of Management and Professor of Marketing and Economics at the SC Johnson College of Business, Cornell University. We are especially grateful to Johannes Habel and Nathaniel Hartmann for their feedback. We thank an anonymous medical equipment firm for collaborating on the field experiment and providing the data. This study was approved by the University of Utah Institutional Review Board (IRB 00179199). The corresponding authors of this paper are Hongye Sun (hongye.sun@eccles.utah.edu) and Kun Peng (pengkun@smail.nju.edu.cn).

we ask a distinct question: does training technology affect workforce selection, even when the provision of training and the skill being taught are held fixed?

We conduct a field experiment with new salesforce recruits at a leading Chinese medical equipment manufacturer that sells to hospitals through a relationship-based, multistage sales process. Salespeople initially contact hospitals to identify physicians and administrators with purchasing authority and obtain their direct contact information for follow-up. Given the length of the sales process, the firm evaluates recruits primarily on the number of sales calls they make, rather than completed sales. For the first five days after hire, recruits receive eight hours of classroom instruction and practice obtaining prospects' direct contact information for two hours per day.

We randomly assign 187 recruits within 12 weekly recruitment cohorts to practice with AI (i.e., the treatment group) or to practice with peers (i.e., the control group). When practicing with peers, recruits interact with others from the same cohort under instructor supervision and receive instructor feedback. When practicing with AI, recruits interact with simulated physicians and administrators powered by a large language model. They choose among simulated prospects of varying difficulty and practice in private with no explicit feedback. Both groups receive the same classroom instruction and are allocated the same amount of practice time.

We study workforce selection on personality traits, motivated by research on individual differences in sales performance [Gong et al., 2024; Chaker et al., 2024; Habel et al., 2024]. During hiring, we measure recruits' Big Five personality traits using a questionnaire [Soto and John, 2017; Zhang et al., 2022]. Given the longstanding belief that productive salespeople are likely to be extraverted [Barrick and Mount, 1991], we focus on extraversion when characterizing the personality composition of the retained workforce. We also measure aggregate sales activity using cumulative sales calls per recruit and track recruits' retention and cumulative employment duration.

We find that practicing with AI rather than practicing with peers causally changes the personality composition of the retained workforce. Two weeks after training completion, average extraversion among retained recruits is 0.76 points lower for those practicing with AI than for those practicing with peers, after adjusting for differences at hire. Despite retaining a less extraverted workforce, practicing with AI increases cumulative sales calls by 53.2 per recruit (+29.3%) from hire through two weeks after training completion. Practicing with AI also increases retention by 13.1 percentage points (+32.2%) and cumulative employment duration by 1.76 days per recruit over the same two-week horizon.

A statistical decomposition [Blundell et al., 2011] attributes 83.8–93.1% of the increase in cumulative sales calls per recruit to additional days of employment among introverts (recruits with extraversion at or below the sample median), partly offset by fewer days of employment among extraverts (-19.1% to -32.1%). Overall, differences in employment duration account for 61.0–64.7% of the increase in cumulative sales calls per recruit. These findings suggest that longer employment among introverts accounts for much of the sales call gains from practicing with AI.

We also find that the retention gains from practicing with AI are larger in larger recruitment cohorts. This may reflect differences in how the training technologies scale with cohort size, po-

tentially arising from differences in instructor attention and the social environment of training [Murthy et al., 2008; Wiseman et al., 2022]. Motivated by this heterogeneity, we conduct an exploratory targeting exercise that assigns training technologies based on recruits' extraversion and recruitment cohort size. This targeted assignment yields more estimated calls per recruit than assigning all recruits to practice with AI or all recruits to practice with peers, although its incremental gain over universal assignment to practice with AI is imprecisely estimated.

Our paper contributes to research on firm-sponsored training and worker selection. This literature studies training as an investment in human capital [Becker, 1964; Acemoglu and Pischke, 1998, 1999], a means of screening workers by ability [Autor, 2001], and a determinant of productivity and retention [Adhvaryu et al., 2023]. We extend this literature by showing that the technology used to deliver training can affect post-hire selection on personality traits, even when the provision of training and the skill being practiced are held fixed.

We also contribute to research on technology-enabled salesforce training. Studies of simulation training, selective training, onboarding, AI coaching, and AI roleplay examine effects on salesperson performance [Murthy et al., 2008; Atefi et al., 2018; Wiseman et al., 2022; Luo et al., 2021; Habel et al., 2026], while related structural work considers training alongside employee selection [Chung et al., 2021]. We extend this research by explicitly examining retention and workforce composition, and quantifying how changes in employment contribute to sales activity per original recruit.

Finally, our study relates to research on human–AI interactions in the workplace [Brynjolfsson et al., 2025]. We study human–AI interactions during training, comparing practicing with AI to practicing with peers. Our findings suggest that workforce selection is an important consideration in evaluating the aggregate consequences of introducing AI into the workplace.

2 Experimental Setting and Data

2.1 Research Setting

We conduct our field experiment at a leading Chinese manufacturer of medical imaging equipment. Its products include mobile C-arm X-ray systems, digital radiography systems, and dental CT scanners, priced from RMB 10,000 to over RMB 1 million. The firm employs approximately 600 salespeople nationwide. Sales involve a multistage process in which salespeople initially contact hospitals to identify physicians and administrators with purchasing authority and obtain their direct contact information for follow-up, which may lead to on-site visits and formal procurement contracts. Given the length of this process, the firm evaluates salespeople primarily on sales call volume, rather than completed sales.

The firm recruits new salespeople in weekly cohorts. During their first five days of employment, recruits receive eight hours of classroom instruction and two hours of supervised sales practice per day. Classroom instruction covers firm-specific product knowledge and local regulations, while practice focuses on obtaining prospects' direct contact information. In the firm's conventional practice format, recruits practice with peers from the same cohort, taking turns playing the

salesperson and the prospective customer under instructor supervision and feedback.

2.2 Experiment Design

The experiment includes 187 recruits from 12 weekly recruitment cohorts. Within each cohort, we randomly assign recruits to practice with AI or to practice with peers. Recruits practicing with peers follow the firm's conventional practice format, with instructor supervision and feedback. Recruits practicing with AI instead interact privately with simulated physicians and hospital administrators powered by a large language model, without instructor supervision or explicit feedback. In both formats, recruits practice obtaining prospective customers' direct contact information for follow-up. All recruits receive the same classroom instruction and are allocated two hours per day for practice over five days. Recruits practicing with AI record an average of 1.85 hours of active interaction[1] with simulated prospects, with five recruits recording no interactions. We do not observe comparable active practice time for recruits practicing with peers.

The AI practice environment consists of 35 simulated prospects that we develop in collaboration with the firm's sales managers, drawing on training materials and historical sales calls. Each prospect represents a prospective customer with distinct characteristics, including their role, equipment needs, and personality, which we specify through prompts. We design 17 prospects to be more difficult by making them less willing to share their direct contact information. Recruits can view each prospect's profile and difficulty before choosing whom to practice with, but cannot see the underlying prompts. They interact with prospects through a text- and voice-based interface and can conduct repeated conversations with the same or different prospects. Web Appendix A details the development and implementation of the simulated prospects.

2.3 Data

To examine how training technology affects workforce selection, we collect recruits' Big Five personality assessments during hiring through the firm's outsourced recruitment agency. The agency measures personality on ten-point scales.[2] The firm uses these scores in hiring and explicitly prefers extraverts, although extraversion is not a formal hiring requirement. This preference, together with prior evidence linking extraversion to sales performance [Barrick and Mount, 1991; Vinchur et al., 1998; Chaker et al., 2024], motivates our focus on extraversion. We link these pre-hire measures to administrative records on treatment assignment, recruitment cohort, hiring and departure dates, starting salary, age, gender, education, department, and subsequent sales activity.

Table 1 reports recruit characteristics and experimental balance. Recruits are predominantly male (95%), 32 years old on average, and typically hold a bachelor's degree. Consistent with the

[1] We measure active interaction time with simulated prospects using message timestamps, excluding intervals longer than 10 minutes to avoid counting inactivity. Using cutoffs between 5 and 60 minutes yields similar estimates (1.61–1.99 hours per recruit).

[2] The assessment uses the Chinese version of the 60-item Big Five Inventory-2 (BFI-2; Soto and John, 2017; Zhang et al., 2022). Each domain comprises 12 items rated on a five-point agreement scale. Domain scores are calculated by averaging responses after reverse-coding negatively keyed items and multiplying by two.

Table 1: Recruit characteristics and experimental balance

Variable	Full Sample	Treatment	Control	Balance p-value
Demographic and employment characteristics				
Male	0.95 (0.21)	0.96 (0.20)	0.94 (0.23)	0.673
Age (years)	31.89 (3.82)	32.02 (4.07)	31.75 (3.54)	0.763
Bachelor's degree or above	0.93 (0.26)	0.92 (0.28)	0.93 (0.25)	0.685
Starting salary (1,000 RMB)	8.39 (1.00)	8.34 (1.03)	8.46 (0.96)	0.547
Cohort size	17.05 (4.76)	17.01 (4.72)	17.10 (4.83)	—
Pre-hire personality traits (1–10)				
Conscientiousness	6.55 (2.09)	6.37 (2.07)	6.75 (2.10)	0.180
Neuroticism	4.93 (2.05)	5.21 (2.00)	4.61 (2.06)	0.107
Openness	5.80 (2.16)	5.79 (2.07)	5.81 (2.26)	0.982
Extraversion	6.12 (2.34)	5.92 (2.20)	6.35 (2.47)	0.175
Agreeableness	5.87 (1.98)	5.68 (1.96)	6.07 (2.00)	0.202
N (recruits)	187	98	89	—

Notes: Means are reported with standard deviations in parentheses. Balance reports two-sided p-values from wild cluster bootstrap tests of stratified differences-in-means, clustered by cohort.

Table 2: Correlations among recruit characteristics at hire

	Pre-Hire Personality Traits					Other Characteristics				
	P1	P2	P3	P4	P5	O1	O2	O3	O4	O5
P1 Conscientiousness	1.00	-0.56^{***}	0.31^{***}	0.47^{***}	0.43^{***}	-0.11	0.14^{*}	-0.09	-0.01	0.04
P2 Neuroticism	-0.56^{***}	1.00	-0.31^{***}	-0.49^{***}	-0.37^{***}	0.07	-0.03	-0.03	0.01	-0.17^{**}
P3 Openness	0.31^{***}	-0.31^{***}	1.00	0.40^{***}	0.08	-0.06	-0.02	-0.09	-0.07	0.10
P4 Extraversion	0.47^{***}	-0.49^{***}	0.40^{***}	1.00	0.32^{***}	-0.10	0.07	0.12	0.04	0.03
P5 Agreeableness	0.43^{***}	-0.37^{***}	0.08	0.32^{***}	1.00	-0.20^{***}	0.07	0.02	0.05	0.04
O1 Age (years)	-0.11	0.07	-0.06	-0.10	-0.20^{***}	1.00	-0.22^{***}	0.06	0.35^{***}	-0.17^{**}
O2 Male	0.14^{*}	-0.03	-0.02	0.07	0.07	-0.22^{***}	1.00	-0.06	0.01	0.13^{*}
O3 Bachelor's or higher	-0.09	-0.03	-0.09	0.12	0.02	0.06	-0.06	1.00	0.07	0.03
O4 Starting salary	-0.01	0.01	-0.07	0.04	0.05	0.35^{***}	0.01	0.07	1.00	-0.07
O5 Cohort size	0.04	-0.17^{**}	0.10	0.03	0.04	-0.17^{**}	0.13^{*}	0.03	-0.07	1.00

Notes: Pearson correlations. Male and bachelor's or higher are binary indicators. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

firm's stated preference for extraverts, recruits score above the midpoint on extraversion on average. We assess experimental balance using the stratified difference-in-means estimator described in Section 3.1. None of the reported differences is statistically significant at the 10% level. Table 2 reports correlations among recruit characteristics measured at hire. Personality traits are moderately correlated with each other but not with starting salary.

We focus on the first two weeks after training completion, when turnover is most concentrated and subsequent on-the-job experiences are less likely to influence outcomes. We characterize workforce composition using the extraversion scores of recruits who remain employed beyond this horizon. We measure retention as remaining employed beyond this horizon and employment duration as the number of days employed from hire through the same horizon. To assess aggregate salesforce activity, we measure cumulative sales calls per recruit from hire through two weeks after training completion, retaining recruits in the sample after they leave employment and

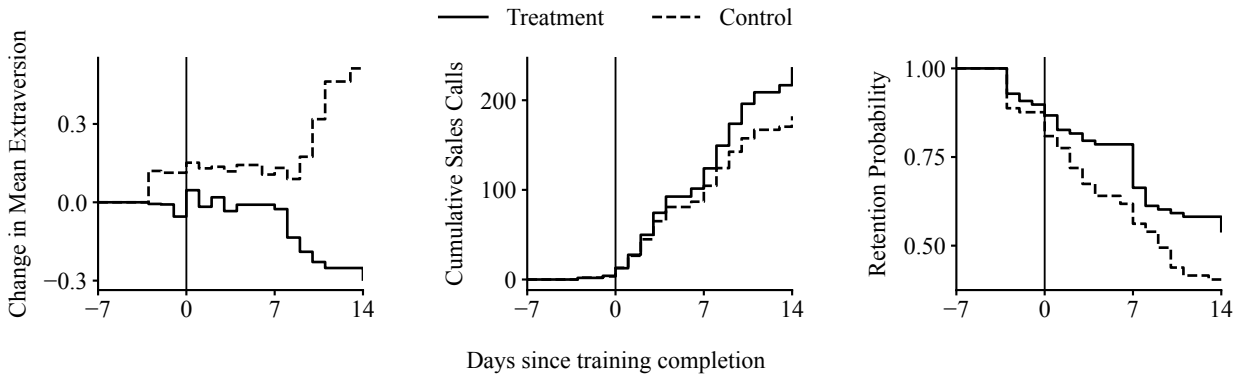
assigning zero calls to these recruits thereafter.

In Web Appendix B, we consider two additional measures of sales success. The firm evaluates salespeople's success in obtaining verified purchase intentions on a calendar-month basis. We define *documented success* as the number of calendar months with at least one verified purchase intention. These evaluations are missing for 70.6% of employed worker-months. To supplement these incomplete records, we construct a second measure from call transcripts: the cumulative number of calls in which salespeople obtain prospective customers' direct contact information for follow-up. We refer to this outcome as *measured success*.

3 Results

Figure 1 plots the evolution of workforce composition, cumulative sales calls per recruit, and retention from hire through two weeks after training completion, separately for recruits practicing with AI and practicing with peers. Workforce composition in the left panel is measured as mean extraversion among recruits still employed on each day, relative to its mean among all recruits assigned to the same training technology. The remaining panels plot cumulative sales calls per recruit and retention probabilities.

Figure 1: Extraversion, cumulative sales calls, and retention probability



The trajectories reveal different patterns across training technologies. Cumulative sales calls per recruit grow more rapidly and retention declines more slowly among recruits practicing with AI. In this section, we estimate the causal effects underlying these descriptive patterns and assess their statistical significance.

3.1 Empirical Strategy

We now estimate the causal effects of practicing with AI rather than practicing with peers on workforce composition, sales activity, and retention (defined as remaining employed beyond two weeks after training completion). For each outcome, let $Y_{ic}(a)$ denote the potential outcome for recruit i in cohort c under assignment a, where $a = 1$ denotes practicing with AI and $a = 0$

denotes practicing with peers. Our intent-to-treat (ITT) estimand is the average treatment effect per assigned recruit across the firm's recruitment cohorts:

$$\tau_Y = \frac{\mathbb{E}\left[\sum_{i=1}^{n_c} \left(Y_{ic}(1) - Y_{ic}(0)\right)\right]}{\mathbb{E}[n_c]}, \tag{1}$$

where n_c denotes cohort size and expectations are over the population of recruitment cohorts. This estimand measures the additional sales calls or days of employment per recruit, or the increase in the share of recruits retained, caused by practicing with AI.

We estimate this effect using the stratified difference-in-means estimator described by Athey and Imbens [2016, Section 6.1], which can equivalently be expressed as a weighted regression with cohort fixed effects:

$$Y_{ic} = \alpha_c + \tau_Y T_{ic} + \varepsilon_{ic}, \tag{2}$$

where T_{ic} indicates assignment to practice with AI and α_c denotes cohort fixed effects. To make the regression coefficient equal the stratified difference-in-means estimator, we weight observations by $1/[p_c(1-p_c)]$, where p_c is the fraction of recruits assigned to practice with AI in cohort c.

For workforce composition, let X_{ic} denote the extraversion of recruit i in cohort c, and let $S_{ic}(a)$ indicate whether that recruit remains employed beyond two weeks after training completion under assignment a. Our estimand is the difference in mean extraversion between the workforces retained under the two training technology assignments:

$$\Delta = \frac{\mathbb{E}[\sum_i X_{ic} S_{ic}(1)]}{\mathbb{E}[\sum_i S_{ic}(1)]} - \frac{\mathbb{E}[\sum_i X_{ic} S_{ic}(0)]}{\mathbb{E}[\sum_i S_{ic}(0)]}. \tag{3}$$

This estimand captures how training technology changes the personality composition of the retained workforce, as measured two weeks after training completion.

To adjust for baseline differences in extraversion between assignment groups, we subtract the difference in mean extraversion between assignment groups at hire. Let $\hat{\mu}_a^R$ and $\hat{\mu}_a^B$ denote the estimated mean extraversion of the retained workforce and all assigned recruits, respectively. Our baseline-adjusted estimator is:

$$\hat{\Delta}^{\text{adj}} = (\hat{\mu}_1^R - \hat{\mu}_1^B) - (\hat{\mu}_0^R - \hat{\mu}_0^B). \tag{4}$$

We implement this estimator using a weighted difference-in-differences regression comparing mean extraversion between recruits practicing with AI and those practicing with peers at hire and two weeks after training completion. The dependent variable in both periods is each recruit's extraversion score. We estimate:

$$X_{ict} = \alpha + \beta T_{ic} + \gamma \text{Post}_t + \delta(T_{ic} \times \text{Post}_t) + \varepsilon_{ict}, \tag{5}$$

where Post_t indicates the observation two weeks after training completion and $X_{ict} = X_{ic}$ for each observed recruit-period. Weighting observations by $1/p_c$ for recruits practicing with AI and $1/(1-$

p_c) for recruits practicing with peers yields an interaction coefficient $\hat{\delta}$ that equals the baseline-adjusted estimator in equation (4).

To account for sampling uncertainty across cohorts in the firm's recruitment process, we cluster by recruitment cohort [Abadie et al., 2023]. Given the 12 observed cohorts, we assess statistical significance using a restricted wild cluster bootstrap [Roodman et al., 2019]. We report asymptotic cluster-robust standard errors and p-values for reference.

3.2 Effect on Workforce Selection

Table 3 shows that practicing with AI changes the personality composition of the retained workforce relative to practicing with peers. Two weeks after training completion, mean extraversion among retained recruits is 0.76 points lower for those practicing with AI, after adjusting for baseline differences. Despite the firm's stated preference for extraverts, practicing with AI produces a retained workforce with lower mean extraversion.

Table 3: Effect of practicing with AI on workforce composition

	Extraversion (1)
Estimated treatment effect	-0.76^{***}
	(0.28)
Asymptotic cluster-robust p-value	0.021
Control change from baseline	0.48
N (recruits)	187

Notes. Baseline-adjusted effect estimated using equation (4). The asymptotic standard error in parentheses is clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

A necessary condition for training technology to change workforce composition is that its proportional effect on retention varies with extraversion.[3] Our finding that practicing with AI produces a retained workforce with lower mean extraversion than practicing with peers thus implies heterogeneity in proportional retention effects by extraversion.

3.3 Effect on Sales Activity

Practicing with AI retains a less extraverted workforce, but generates more aggregate sales activity from the original pool of recruits. Table 4 shows that practicing with AI increases cumulative sales calls by 53.2 per recruit, a 29.3% increase relative to the control mean of 181.5 calls. In Web Appendix B, we also find that practicing with AI increases measured success by 4.19 (+62.9%) calls per recruit. The estimated effect on documented success is positive but imprecisely estimated.

[3]Consider recruits with high (H) and low (L) levels of a pre-hire personality trait. Let $p_g(a)$ denote the retention probability for group $g \in \{H, L\}$ under treatment ($a = 1$) or control ($a = 0$). Workforce composition changes if and only if $p_H(1)/p_L(1) \neq p_H(0)/p_L(0)$. This condition requires heterogeneity in proportional treatment effects on retention.

Table 4: Effect of practicing with AI on sales calls per recruit

	Sales calls per recruit (1)
Estimated treatment effect	53.21^{**}
	(21.85)
Asymptotic cluster-robust p-value	0.033
Control mean	181.47
N (recruits)	187

Notes. Intent-to-treat effect estimated using equation (2). The asymptotic standard error in parentheses is clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

3.4 Effect on Retention

Prior research on AI-enabled sales training emphasizes human capital formation as a source of performance gains [Luo et al., 2021; Habel et al., 2026]. However, aggregate sales activity also depends on how long recruits remain employed. We examine whether practicing with AI increases retention and how additional employment accounts for the increase in calls per recruit.

Table 5 shows that practicing with AI increases retention by 13.1 percentage points, a 32.2% increase relative to the control mean of 40.7%. Practicing with AI also increases cumulative employment duration from hire through two weeks after training completion by 1.76 days per recruit.

Table 5: Effects of practicing with AI on retention and employment duration

	Retention (1)	Employment duration (days) (2)
Estimated treatment effect	0.131^{**}	1.76^{***}
	(0.051)	(0.55)
Asymptotic cluster-robust p-value	0.026	0.009
Control mean	0.407	14.82
N (recruits)	187	187

Notes. Intent-to-treat effects estimated using equation (2). The asymptotic standard errors in parentheses are clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

3.4.1 Accounting for Sales Call Gains by Extraversion

To assess how changes in employment contribute to the increase in calls per recruit, we conduct a statistical decomposition following Blundell et al. [2011]. We separate the contributions of employment duration and calling activity while employed, and further distinguish the employment contributions of introverts and extraverts, defined by whether their extraversion scores are at or below the sample median.

For each cohort, day, and personality group, let s_a denote the fraction of all recruits assigned to arm a who belong to that group and remain employed, and let r_a denote the average number of calls made that day by employed recruits in that group. Their product, $s_a r_a$, gives the group's

contribution to sales calls per originally assigned recruit.

We decompose the difference in calls between practicing with AI ($a = 1$) and practicing with peers ($a = 0$) into two components: a contribution from changes in employment, holding calling rates fixed, and a contribution from changes in calling rates among employed recruits. There are two ways to make this calculation:

$$s_1 r_1 - s_0 r_0 = \underbrace{(s_1 - s_0) r_0}_{\text{Employment contribution}} + \underbrace{s_1 (r_1 - r_0)}_{\text{Calling-rate contribution}}$$

$$= \underbrace{(s_1 - s_0) r_1}_{\text{Employment contribution}} + \underbrace{s_0 (r_1 - r_0)}_{\text{Calling-rate contribution}} .$$

The first calculation values additional employment using the calling rate under peer practice; the second uses the calling rate under AI practice. In each case, the remaining difference is attributed to calling activity while employed. We sum these contributions across days and personality groups, weighting cohorts by their shares of the original sample.

Table 6 shows that additional employment among introverts accounts for 83.8–93.1% of the increase in calls per recruit, partly offset by reduced employment among extraverts. Thus, despite the firm's stated preference for extraverts, additional employment among introverts is the largest contributor to the sales call gains from practicing with AI.

Table 6: Statistical decomposition of the increase in cumulative sales calls per recruit

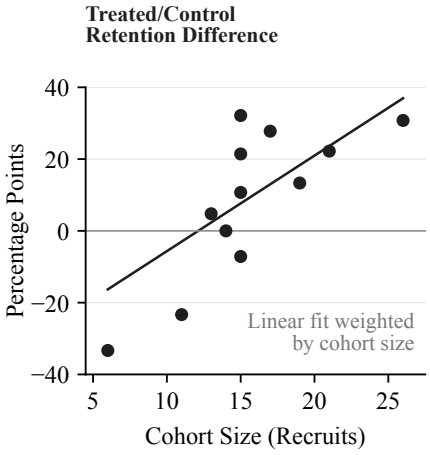
	Additional calls per recruit	Share of total increase
Changes in employment		
Additional employment among introverts	$+44.6$ – $+49.5$	$+83.8\%$ – $+93.1\%$
Reduced employment among extraverts	-17.1 – -10.2	-32.1% – -19.1%
Net employment contribution	$+32.5$ – $+34.4$	$+61.0\%$ – $+64.7\%$
Changes in calling activity while employed	$+18.8$ – $+20.7$	$+35.3\%$ – $+39.0\%$
Total increase in sales calls	+53.2	100%

3.4.2 Heterogeneity with Cohort Size

We examine whether the effect of practicing with AI on retention varies with recruitment cohort size. The left panel of Figure 2 plots the treatment–control difference in retention against cohort size. The right panel reports the interaction between treatment assignment and continuous cohort size, estimated by augmenting equation (2) with this interaction.

The estimated retention gain from practicing with AI increases by 2.7 percentage points per additional recruit. This heterogeneity may reflect differences in how the training technologies scale with cohort size. Larger cohorts may limit instructor attention and increase exposure to peers during practice [Murthy et al., 2008; Wiseman et al., 2022]. Our limited sample size prevents us from distinguishing between these explanations.

Figure 2: Cohort size and retention



	Retention (1)
Treatment × cohort size	0.027**
	(0.006)
Asymptotic cluster-robust p-value	0.002
Control mean	0.407
N (recruits)	187

Notes. Equation (2) augmented with a treatment × cohort-size interaction. The asymptotic standard error in parentheses is clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

3.4.3 Targeting Training Technologies by Personality and Cohort Size

Differences in retention by extraversion and cohort size motivate an exploratory exercise in assigning training technologies based on these characteristics. We consider a rule under which recruits practice with AI if their extraversion scores are at or below the sample median, or they belong to cohorts larger than the sample median. Only extraverts in smaller cohorts practice with peers.

Using outcomes from the randomized experiment, we estimate that this rule generates 242.9 cumulative calls per recruit, compared with 234.7 when all recruits practice with AI and 181.5 when all recruits practice with peers. The incremental gain of 8.2 calls per recruit over assigning all recruits to practice with AI is imprecisely estimated. These results motivate further testing of whether targeting training technologies can improve aggregate sales activity.

3.5 Treatment Effect Dynamics

To examine treatment effect dynamics, we reestimate the effects on our main outcomes from one through five weeks after training completion. Table 7 shows that the difference in mean extraversion among retained recruits is small at one week but grows in magnitude at each subsequent horizon, reaching -1.41 points at five weeks. The gains in cumulative sales calls and employment duration also increase with the observation window, reaching 113.9 additional calls and 3.70 additional days per recruit at five weeks. The retention advantage is largest at two weeks, when turnover is concentrated shortly after training completion, and declines from 13.1 percentage points at that horizon to 6.9 percentage points at five weeks. These dynamics suggest that differences in workforce composition develop over time and persist beyond the primary two-week horizon, whereas turnover is concentrated at two weeks after training completion.

Table 7: Treatment effect dynamics

	Weeks since training completion				
	1	2	3	4	5
Extraversion	-0.12	-0.76^{***}	-1.11^{**}	-1.28^{**}	-1.41^{**}
	(0.19)	(0.28)	(0.52)	(0.61)	(0.68)
Sales calls per recruit	19.10^{*}	53.21^{**}	75.96^{**}	97.41^{**}	113.88^{**}
	(11.67)	(21.85)	(33.08)	(42.96)	(52.39)
Employment duration (days)	0.84^{***}	1.76^{***}	2.54^{**}	3.23^{**}	3.70^{**}
	(0.27)	(0.55)	(0.98)	(1.33)	(1.62)
Retention	0.098^{*}	0.131^{**}	0.102	0.065	0.069
	(0.046)	(0.051)	(0.078)	(0.048)	(0.048)
N (recruits)	187	187	187	187	187

Notes. The effect on extraversion is estimated using equation (4). Intent-to-treat effects on the remaining outcomes are estimated using equation (2). Sales calls and employment duration are cumulative from hire through each horizon; retention is remaining employed beyond each horizon. The asymptotic standard errors in parentheses are clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

3.6 Additional Analyses and Robustness Checks

We examine workforce composition across all five pre-hire personality traits. In addition to extraversion, practicing with AI produces a retained workforce with lower mean conscientiousness (-0.82 points). Both effects remain statistically significant after Holm correction across the five traits (adjusted $p < 0.01$ for both). The effects on neuroticism, openness, and agreeableness do not survive this correction. However, conscientiousness and extraversion are positively correlated at hire ($r = 0.47$, $p < 0.01$), so these findings do not necessarily reflect distinct patterns.

To examine whether the estimated effects on our main outcomes are driven by individual cohorts, we reestimate each effect after excluding one of the 12 cohorts in turn. The estimated effects on workforce composition, cumulative sales calls per recruit, employment duration, and retention retain their original signs across all exclusions, with magnitudes ranging from 0.58 to 0.90 extraversion points, 33.5 to 58.6 calls per recruit, 1.51 to 2.05 days per recruit, and 10.3 to 15.4 percentage points, respectively. These results suggest that our findings are not driven by any single cohort.

4 Conclusion

Our field experiment shows that the choice of training technology can change workforce composition after hiring. Practicing with AI produces a retained workforce with lower mean extraversion. Yet practicing with AI increases cumulative sales calls per recruit by 29.3%. A statistical decomposition shows that additional days of employment among introverts account for 83.8–93.1% of this increase, partly offset by fewer days of employment among extraverts.

For firms, these findings suggest evaluating hiring and training decisions jointly. The characteristics firms favor at hiring may have different implications for retention and aggregate perfor-

mance depending on the training technology used. Firms may improve aggregate performance by choosing training technologies that retain workers whose characteristics differ from those traditionally favored at hiring. The larger retention gains from practicing with AI in larger cohorts also motivate testing whether tailoring training technologies to recruit characteristics and cohort size can further improve aggregate performance.

Workforce selection may also have consequences beyond the sales outcomes we measure. Training technologies can alter the distribution of personality traits among workers who subsequently interact with colleagues, managers, and customers. Research in organizational behavior links group personality composition to cooperation, helping behavior, and team performance [Barry and Stewart, 1997; Beersma et al., 2003; Gonzalez-Mulé et al., 2014]. AI is increasingly being integrated into training across domains, particularly in education [Bastani et al., 2025; Kestin et al., 2025] and surgical training [Fazlollahi et al., 2022; Giglio et al., 2025]. As these technologies reshape how trainees develop skills, our findings motivate examining their effects on participation, retention, and workforce composition across settings.

Our study has several limitations. The experiment covers 187 recruits at one medical equipment firm, and our primary analysis focuses on outcomes through two weeks after training completion, with additional estimates extending through five weeks. Our measures capture sales activity and progress through the sales process, leaving longer-run effects on revenue and customer relationships open. The training technologies also differ in instructor supervision, peer interactions, and opportunities to choose simulated prospects, limiting our ability to isolate the mechanisms underlying retention. Future research can examine whether similar workforce selection effects arise across training technologies, occupations, and longer employment horizons.

Declaration of Funding and Competing Interests. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

References

Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. When Should You Adjust Standard Errors for Clustering? *The Quarterly Journal of Economics*, 138(1):1–35, 2023.

Daron Acemoglu and Jörn-Steffen Pischke. Why Do Firms Train? Theory and Evidence. *The Quarterly Journal of Economics*, 113(1):79–119, 1998.

Daron Acemoglu and Jörn-Steffen Pischke. The Structure of Wages and Investment in General Training. *Journal of Political Economy*, 107(3):539–572, 1999.

Achyuta Adhvaryu, Namrata Kala, and Anant Nyshadham. Returns to On-the-Job Soft Skills Training. *Journal of Political Economy*, 131(8):2165–2208, 2023.

Mark Armstrong. How Microsoft Built an AI Coach to Scale Personalized Sales Training. Microsoft Inside Track, 2026.

Yashar Atefi, Michael Ahearne, James G. Maxham, III, D. Todd Donavan, and Brad D. Carlson. Does Selective Sales Force Training Work? *Journal of Marketing Research*, 55(5):722–737, 2018.

Susan Athey and Guido W. Imbens. The Econometrics of Randomized Experiments. arXiv:1607.00698, 2016.

David H. Autor. Why Do Temporary Help Firms Provide Free General Skills Training? *The Quarterly Journal of Economics*, 116(4):1409–1448, 2001.

Murray R. Barrick and Michael K. Mount. The Big Five Personality Dimensions and Job Performance: A Meta-Analysis. *Personnel Psychology*, 44(1):1–26, 1991.

John M. Barron, Dan A. Black, and Mark A. Loewenstein. Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth. *Journal of Labor Economics*, 5(1):76–89, 1987.

Bruce Barry and Greg L. Stewart. Composition, Process, and Performance in Self-Managed Groups: The Role of Personality. *Journal of Applied Psychology*, 82(1):62–78, 1997.

Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26):e2422633122, 2025.

Gary S. Becker. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. National Bureau of Economic Research, New York, 1964.

Bianca Beersma, John R. Hollenbeck, Stephen E. Humphrey, Henry Moon, Donald E. Conlon, and Daniel R. Ilgen. Cooperation, Competition, and Team Performance: Toward a Contingency Approach. *Academy of Management Journal*, 46(5):572–590, 2003.

Richard Blundell, Antoine Bozio, and Guy Laroque. Labor Supply and the Extensive Margin. *American Economic Review*, 101(3):482–486, 2011.

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative AI at Work. *The Quarterly Journal of Economics*, 2025.

Nawar N. Chaker, Johannes Habel, Nathaniel N. Hartmann, Felix Johannsen, and Heinrich Rusche. Quiet sellers: When introversion drives salesperson performance. *Journal of Retailing*, 100(3):456–474, 2024.

Doug J Chung, Byungyeon Kim, and Byoung G Park. The Comprehensive Effects of Sales Force Management: A Dynamic Structural Analysis of Selection, Compensation, and Training. *Management Science*, 67(11):7046–7074, 2021.

Jared Council. MetLife Says AI Is Improving Its Call Centers. The Wall Street Journal, 2019.

Ali M. Fazlollahi, Mohamad Bakhaidar, Ahmad Alsayegh, Recai Yilmaz, Alexander Winkler-Schwartz, Nykan Mirchi, Ian Langleben, Nicole Ledwos, Abdulrahman J. Sabbagh, Khalid Bajunaid, Jason M. Harley, and Rolando F. Del Maestro. Effect of Artificial Intelligence Tutoring vs Expert Instruction on Learning Simulated Surgical Skills Among Medical Students: A Randomized Clinical Trial. *JAMA Network Open*, 5(2), 2022.

Bianca Giglio, Abdulmajeed Albeloushi, Ahmad Kh Alhaj, Mohamed Alhantoobi, Rothaina Saeedi, Vanja Davidovic, Abicumaran Uthamacumaran, Recai Yilmaz, Jason Lapointe, Neevya Balasubramaniam, Trisha Tee, Ali M. Fazlollahi, José A. Correa, and Rolando F. Del Maestro. Artificial Intelligence-Augmented Human Instruction and Surgical Simulation Performance: A Randomized Clinical Trial. *JAMA Surgery*, 160(9):993–1003, 2025.

Shiyang Gong, Qian Li, Song Su, and Juanjuan Zhang. Genes and Sales. *Management Science*, 70(6):3902–3922, 2024.

Gabriel R. Gonzalez, Johannes Habel, and Gary K. Hunter. AI Agents, Agentic AI, and the Future of Sales. *Journal of Business Research*, 2026.

Erik Gonzalez-Mulé, David S. DeGeest, Brian W. McCormick, Jee Young Seong, and Kenneth G. Brown. Can We Get Some Cooperation Around Here? The Mediating Role of Group Norms on the Relationship Between Team Personality and Individual Helping Behaviors. *Journal of Applied Psychology*, 99(5):988–999, 2014.

Johannes Habel, Michael Ahearne, Seshadri Tirunillai, and Amy Vandaveer Novak. Do AI Role Plays Improve Salesperson Performance? Evidence from a Longitudinal Field Study. Technical report, SSRN, 2026. Working paper, revised September 30, 2026.

Johannes Habel, Selma Kadić-Maglajlić, Nathaniel N. Hartmann, Ad de Jong, Nicolas A. Zacharias, and Fabian Kosse. Neuroticism and the sales profession. *Organizational Behavior and Human Decision Processes*, 184, 2024.

Boyan Jovanovic. Firm-Specific Capital and Turnover. *Journal of Political Economy*, 87(6):1246–1260, 1979.

Greg Kestin, Kelly Miller, Anna Klales, Timothy Milbourne, and Gregorio Ponti. AI tutoring outperforms in-class active learning: an RCT introducing a novel research-based design in an authentic educational setting. *Scientific Reports*, 15(1):17458, 2025.

LinkedIn Learning. Workplace Learning Report: The Rise of Career Champions. 2025.

Xueming Luo, Marco Shaojun Qin, Zheng Fang, and Zhe Qu. Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. *Journal of Marketing*, 85(2):14–32, 2021.

Microsoft. Ageas Revolutionizes Sales Training with a New Digital Coach, Powered by Azure OpenAI. 2025.

Nagesh N Murthy, Goutam N Challagalla, Leslie H Vincent, and Tasadduq A Shervani. The Impact of Simulation Training on Call Center Agent Performance: A Field-Based Investigation. *Management Science*, 54(2):384–399, 2008.

Donald O. Parsons. Specific Human Capital: An Application to Quit Rates and Layoff Rates. *Journal of Political Economy*, 80(6):1120–1143, 1972.

David Roodman, Morten Ørregaard Nielsen, James G. MacKinnon, and Matthew D. Webb. Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1):4–60, 2019.

Christopher J. Soto and Oliver P. John. The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model with 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology*, 113(1):117, 2017.

Andrew J. Vinchur, Jeffery S. Schippmann, Fred S. Switzer, III, and Philip L. Roth. A Meta-Analytic Review of Predictors of Job Performance for Salespeople. *Journal of Applied Psychology*, 83(4):586–597, 1998.

Phillip Wiseman, Michael Ahearne, Zachary Hall, and Seshadri Tirunillai. Onboarding Salespeople: Socialization Approaches. *Journal of Marketing*, 86(6):13–31, 2022.

Bo Zhang, Yi Ming Li, Jian Li, Jing Luo, Yonghao Ye, Lu Yin, Zhuosheng Chen, Christopher J. Soto, and Oliver P. John. The Big Five Inventory–2 in China: A Comprehensive Psychometric Evaluation in Four Diverse Samples. *Assessment*, 29(6):1262–1284, 2022.

Web Appendix A: Experiment Implementation

This appendix describes the development of the simulated prospects, the AI practice interface, and recruits' interactions with the AI practice environment.

A.1 Development and Design of Simulated Prospects

We developed simulated prospects in collaboration with the firm's sales managers over the year preceding the experiment. Each simulated prospect is implemented as a system prompt for the Qwen Plus large language model, specifying characteristics such as professional role, demographics, personality, communication style, equipment needs, and willingness to purchase. Our objective was to create prospective customers whose characteristics and conversational behavior reflected those encountered by the firm's salespeople. We drew on the firm's training materials and product manuals to establish the simulated prospects' technical knowledge and historical sales calls to inform their conversational behavior. Development proceeded in three phases.

In the first phase, we constructed 50 simulated prospects using these materials. Sales managers evaluated them and found that they were uniformly unwilling to purchase and lacked realistic personalities. In the second phase, we revised the system prompts to introduce distinct personalities and variation in difficulty. Approximately half of the simulated prospects were designed to be easier to engage, while the remainder were made more reluctant to share their direct contact information. Sales managers evaluated the revised simulated prospects and removed those they considered redundant, leaving 35. In the third phase, we tested the 35 simulated prospects with 20 new hires. These tests revealed that recruits lacked sufficient information to choose which simulated prospects to practice with and did not consistently understand the objective of a conversation. We therefore introduced visible prospect descriptions and difficulty levels and explicitly instructed recruits to obtain the prospect's direct contact information.

The final set consists of 35 simulated prospects, of which 18 are designed to be easier to obtain direct contact information from, and 17 designed to be difficult. Their system prompts specify additional characteristics, including hospital and department, current equipment, supplier relationships, department budgets, patient volume, and facility constraints. These characteristics shape their conversational behavior and the concerns they raise. For example, simulated orthopedic surgeons emphasize equipment performance in surgical workflows, whereas simulated radiologists emphasize imaging reliability and patient throughput. Table A1 lists the simulated prospects' demographic characteristics, professional roles, product needs, and difficulty levels. The system prompts are the firm's intellectual property and are not reproduced verbatim.

A.2 Practice Interface and Task

Recruits interact with simulated prospects through a conversational interface supporting text and voice. Before a conversation, recruits can select from 35 simulated prospects and view a profile containing the prospect's name, image, department, professional role, and a brief description of

Table A1: AI practice personas

Index	Gender	Age	Department	Role	Product Need	Difficulty
1	Male	40	Orthopedics	Physician	Surgical Robot	Easy
2	Male	37	Orthopedics	Attending Physician	C-arm	Easy
3	Male	48	Orthopedics	Orthopedics Director	3D C-arm	Hard
4	Male	35	Orthopedics	Attending Physician	3D C-arm	Easy
5	Male	48	Orthopedics	Orthopedics Director	3D C-arm	Hard
6	Female	35	Orthopedics	Attending Physician	C-arm	Hard
7	Female	38	Orthopedics	Attending Physician	C-arm	Hard
8	Female	35	Radiology	Attending Physician	DR System	Hard
9	Male	38	Orthopedics	Attending Physician	C-arm	Easy
10	Female	38	Radiology	Attending Physician	Fixed DR	Easy
11	Male	40	Orthopedics	Department Director	C-arm	Easy
12	Male	37	Orthopedics	Attending Physician	3D C-arm	Easy
13	Female	47	Dentistry	Department Director	Dental CT	Easy
14	Female	48	Orthopedics	Department Director	C-arm	Hard
15	Female	48	Gastroenterology	Department Director	Fixed DR	Hard
16	Male	45	Orthopedics	Department Director	C-arm	Easy
17	Male	42	Medical Equipment	Department Director	C-arm	Easy
18	Male	40	Orthopedics	Department Director	C-arm	Hard
19	Female	38	Dentistry	Attending Physician	Dental CT	Easy
20	Male	45	Gastroenterology	Department Director	C-arm	Hard
21	Female	35	Radiology	Attending Physician	Fixed DR	Hard
22	Male	45	Medical Equipment	Purchasing Specialist	C-arm/DR	Hard
23	Male	37	Orthopedics	Attending Physician	C-arm	Easy
24	Female	35	Radiology	Attending Physician	Dynamic DR	Easy
25	Female	42	Radiology	Attending Physician	Dynamic DR	Hard
26	Male	40	Radiology	Hospital Director	DR System	Hard
27	Male	39	Neurosurgery	Attending Physician	C-arm	Easy
28	Male	45	Orthopedics	Attending Physician	3D C-arm	Easy
29	Female	37	Orthopedics	Attending Physician	C-arm	Hard
30	Female	35	Radiology	Attending Physician	Fixed DR	Hard
31	Male	35	Medical Equipment	Purchasing Specialist	C-arm	Easy
32	Male	45	Orthopedics	Orthopedics Director	3D C-arm	Easy
33	Male	48	Orthopedics	Department Director	C-arm	Easy
34	Male	50	Orthopedics	Orthopedics Director	3D C-arm	Hard
35	Male	37	Orthopedics	Attending Physician	C-arm	Hard

their personality, equipment needs, and familiarity with the firm. The interface also displays the prospect's difficulty level. Recruits cannot view the underlying system prompts. Table A2 reproduces the profiles' textual summaries, while Figure A1 illustrates the interface, including the conversation history, instructions, active conversation window, and prospect selection menu.

The practice task mirrors the initial stage of the firm's sales process. In actual sales calls, salespeople generally begin by contacting a hospital's front desk or another publicly available number. Their immediate objective is to identify a physician or administrator with purchasing authority and obtain that person's direct contact information for subsequent follow-up. Accordingly, recruits practicing with AI are instructed to obtain the prospect's direct contact information. A practice conversation succeeds when the prospect provides this information.

Within the allocated practice period, recruits can choose which simulated prospects to engage and conduct repeated conversations with the same or different prospects, with no required min-

imum number of conversations. Each simulated prospect's characteristics remain fixed across conversations, allowing recruits to try different sales strategies with the same prospect.

Simulated prospects serve as practice partners rather than coaches: they respond as prospective customers but do not evaluate recruits' selling techniques or provide advice. Unlike recruits practicing with peers, those practicing with AI receive no instructor supervision or explicit feedback during practice. They can use success in obtaining direct contact information to inform subsequent conversations. Figure A2 illustrates a conversation with a difficult simulated prospect who expresses uncertainty about the product and requests references from other hospitals. Figure A3 shows recruits practicing with simulated prospects.

Table A2: Observable summaries of AI practice personas

Index	Summary
1	Careful and proactive orthopedics physician familiar with the firm's C-arm; the department is considering an orthopedic surgical robot compatible with a two-dimensional small C-arm.
2	Friendly attending orthopedist interested in new technology; the department plans to replace an aging imported C-arm and is interested in ease of operation, imaging speed, and service.
3	Rigorous orthopedics director who saw the firm's three-dimensional C-arm at an equipment expo, liked the image quality, but has concerns about operational complexity.
4	Detail-oriented attending orthopedist satisfied with the firm's PLX118F flat-panel small C-arm; the department is considering a three-dimensional C-arm as surgery volume grows.
5	Pragmatic orthopedics director with doubts about domestic brands; the department uses an aging Philips three-dimensional C-arm and requires strong evidence on stability and service.
6	Detail-oriented attending orthopedist who has tried the firm's PLX118WF wireless C-arm; the department is preparing to update a GE flat-panel small C-arm.
7	Responsible but conservative attending orthopedist who prefers imported brands; the department may update its GE C-arm and wants technical strength and reliable service.
8	Busy radiology attending physician who is open to new technology; the department may replace an eight-year-old Siemens DR system and needs better efficiency and image quality.
9	Technology-oriented attending orthopedist with some prior knowledge of the firm; the department is considering replacing a Philips flat-panel small C-arm.
10	Collaborative radiology attending physician who has heard of the firm; the department plans to update an older fixed DR and cares about image quality, usability, and service.
11	Open and curious orthopedics director with no prior firm contact; he wants market information even though the department has no immediate C-arm purchase plan.
12	Open-minded attending orthopedist impressed by the firm's PLX7500 three-dimensional C-arm at an industry meeting; the department may update an older two-dimensional C-arm.
13	Careful dentistry department director interested in the firm's dental CT; the department may replace an aging dental CT that often fails under heavy use.
14	Cautious orthopedics director who has heard of the firm but has not used its products; the department may update a Philips C-arm and is skeptical of domestic equipment.
15	Demanding gastroenterology director with limited brand knowledge; the department plans to update an old fixed DR and needs strong answers on product performance and service.
16	Rigorous orthopedics director with some familiarity with the firm; the department may update a GE flat-panel small C-arm and values imaging quality and service.

Continued on next page

Table A2: Observable summaries of AI practice personas (continued)

Index	Summary
17	Outgoing equipment department director who understands and somewhat approves of the firm; the hospital is considering a GE small C-arm, so the salesperson must build support through the equipment office.
18	Experienced orthopedics director with strong views and little patience for low-value calls; he knows the firm but is comparing many brands for a possible equipment update.
19	Detail-oriented dentistry attending physician cautiously open to new products; the department may update an old imported dental CT with low resolution.
20	Analytical gastroenterology director focused on reliability and service; the department may upgrade a six-year-old C-arm used in interventional procedures.
21	Demanding radiology attending physician who has heard of the firm but not used its products; the department is considering replacing an old Wandong fixed DR.
22	Equipment department purchasing specialist who evaluates suppliers carefully; the hospital is cautious about domestic brands but may update older C-arm and DR equipment.
23	Easygoing attending orthopedist impressed by the firm's PLX118WF wireless C-arm; the department may replace an aging Philips two-dimensional C-arm.
24	Direct radiology attending physician satisfied with the firm's PLX8600 dynamic DR; she wants to learn about newer models as the department considers a DR upgrade.
25	Data-oriented radiology attending physician who has never used the firm's products and prefers imported brands; the department may update a Siemens dynamic DR.
26	Careful hospital director who is conservative about new technology; the radiology department uses old firm equipment that no longer meets demand and may purchase additional devices.
27	Neurosurgery attending physician who heard positive comments about the firm's PLX7500 from a colleague; the department may replace an aging GE C-arm.
28	Rigorous attending orthopedist impressed by the firm's PLX7500 three-dimensional C-arm; the department may update an older Siemens three-dimensional C-arm.
29	Ambitious attending orthopedist with high expectations for new technology; she has heard about the firm's PLX7500 but remains in an evaluation stage.
30	Careful radiology attending physician who previously used the firm's PLX8600 but was not fully satisfied; she wants to hear about improvements before a fixed DR update.
31	Equipment department purchasing specialist focused on efficiency and cost control; the hospital is evaluating a new C-arm purchase and wants information on the firm's PLX7500.
32	Team-oriented orthopedics director interested in the firm's PLX7500 after a product exhibition; the department may replace an older GE three-dimensional C-arm.
33	Responsible orthopedics director focused on improving surgical success; he has heard about the firm's PLX7500 but has not used it and remains cautious.
34	Careful orthopedics director with high brand standards and no firm usage experience; the department may update a Philips three-dimensional C-arm.
35	Responsible attending orthopedist with some awareness of the firm; the department may update a GE flat-panel small C-arm and wants improved image quality and positioning precision.

A.3 Observed Practice Behavior

We use practice sessions recorded while recruits remain employed to describe their choices of simulated prospects and success in obtaining contact information.

Figure A1: AI practice interface



Notes: The interface comprises 4 components: conversation history (A), instructions and warnings (e.g., "when switching doctors, first click new chat") (B), the active conversation window (C), and the persona selection menu (D).

Figure A4 presents the sequence of practice sessions for each recruit assigned to AI practice, distinguishing between easier and more difficult simulated prospects. Recruits vary in both the number of conversations they conduct and the proportion involving difficult prospects. On average, treated recruits practice for 1.85 hours and conduct 10.64 conversations, with an average share of 51.9% involving difficult prospects. Five recruits have no recorded practice sessions while employed.

Figure A5 examines practice success. Panel (a) reports the proportion of conversations in which recruits obtained direct contact information from each simulated prospect. Prospects designed to be difficult have substantially lower observed success rates than easier prospects. Panel (b) plots each recruit's success rate against the proportion of conversations involving difficult prospects, showing how success varies with prospect selection.

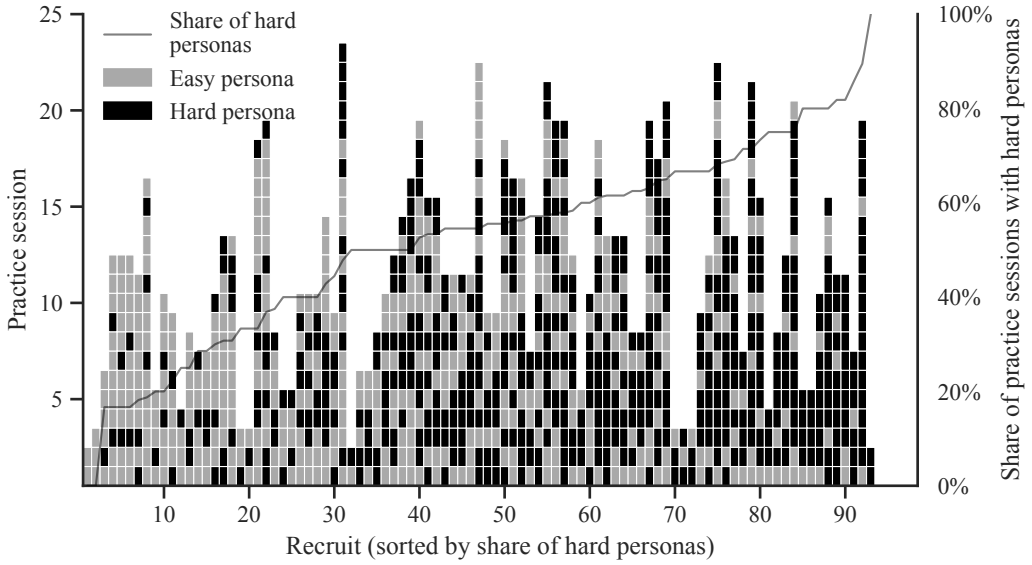
Figure A2: Sample AI practice conversation with a hard persona



Figure A3: Recruits practicing with simulated prospects

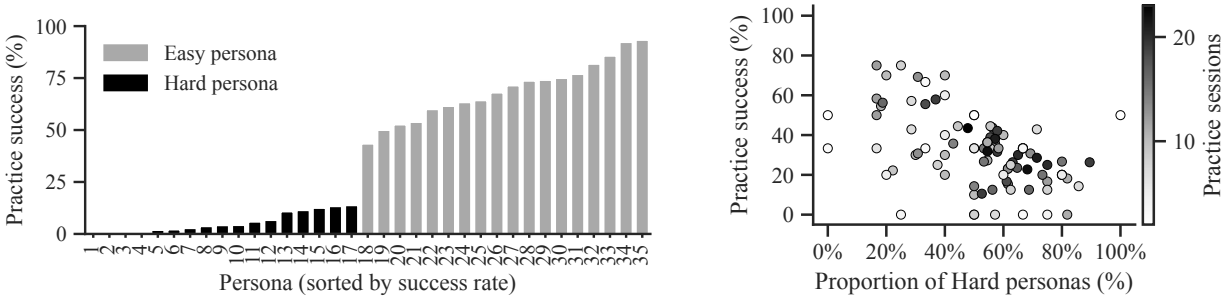


Figure A4: AI practice sequences by recruit



Notes: The figure plots practice sessions for the 98 recruits assigned to treatment. Each column is a recruit, sorted by the share of sessions with hard personas. Empty columns indicate the five recruits with no recorded practice sessions while employed. Each row is a session number within a recruit. Gray cells indicate easy-persona sessions and black cells indicate hard-persona sessions. The gray line plots each recruit's share of sessions with hard personas.

Figure A5: AI practice success by persona and share of hard personas



(a) **(b)**

Notes: Panel (a) plots the success rate for each persona, sorted by success rate. Success means obtaining the persona's direct contact information for follow-up. Panel (b) plots each treated recruit's practice success rate against the share of sessions with hard personas, among recruits with recorded practice sessions.

Web Appendix B: Supplementary Results on Sales Success

We consider two additional measures of sales success. The firm evaluates salespeople's success in obtaining verified purchase intentions on a calendar-month basis. We define *documented success* as the number of calendar months with at least one verified purchase intention, treating months without an evaluation as having no documented success. These evaluations are missing for 70.6% of employed worker-months, and 67.9% of recruits have no evaluation.

To supplement these incomplete records, we construct a second measure from call transcripts: the cumulative number of calls in which salespeople obtain prospective customers' direct contact information for follow-up. We call this *measured success*. We use regular expressions to identify calls in which the salesperson requests direct contact information using terms such as "phone number," "phone," "landline," "contact information," or "how to contact," and the prospect provides a phone number containing seven to eleven digits within the next two conversational turns. We validate this regular expression by manually inspecting a sample of transcripts.

Table B1 shows that practicing with AI increases measured success by 4.19 calls per recruit, relative to a control mean of 6.66. The estimated effect on documented success is also positive, at 0.26 months per recruit relative to a control mean of 0.28, but is imprecisely estimated. These findings show that practicing with AI increases calls yielding direct contact information alongside overall call volume, while the incomplete coverage of the firm's evaluation records limits the conclusions we can draw about verified purchase intentions.

Table B1: Effects of practicing with AI on sales success

	Measured success (1)	Documented success (2)
Estimated treatment effect	4.19^{**}	0.26^{*}
	(1.76)	(0.14)
Asymptotic cluster-robust p-value	0.037	0.094
Control mean	6.66	0.28
N (recruits)	187	187

Notes. Intent-to-treat effects estimated using equation (2). The asymptotic standard errors in parentheses are clustered by cohort. Statistical significance is based on wild cluster bootstrap p-values. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.