

Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions

Xueming Luo

Charles Gilliland Chair Distinguished Professor of Marketing, Strategy, and MIS, and
Founder/Director of the Global Center for Big Data in Mobile Analytics, Fox School of
Business, Temple University Xueming.luo@temple.edu

Marco Shaojun Qin

Assistant Professor of Marketing, Fox School of Business, Temple University
Shaojun.qin@temple.edu

Zheng Fang

Professor of Marketing, Business School, Sichuan University 149281891@qq.com

Zhe Qu

Associate Professor of Information Systems, School of Management, Fudan University
quz@fudan.edu.cn

Journal of Marketing Forthcoming

Sep 2020

*We thank a number of colleagues and seminar participants for feedback and helpful advice. We gratefully acknowledge the anonymous companies for sponsoring the field experiments. All errors and omissions remain our responsibility.

Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions

Abstract

Firms are exploiting AI coaches to provide training to sales agents and improve their job skills. The authors present several caveats associated with such practices based on a series of randomized field experiments. Experiment 1 shows that the incremental benefit of the AI coach over human managers is heterogeneous across agents in an inverted-U shape. While middle-ranked agents improve their performance by the largest amount, both bottom- and top-ranked agents show limited incremental gains. This pattern is driven by a learning-based mechanism in which bottom-ranked agents encounter the most severe information overload problem with the AI versus human coach, while top-ranked agents hold the strongest aversion to the AI relative to human coach. To alleviate the challenge faced by bottom-ranked agents, experiment 2 redesigns the AI coach by restricting the training feedback level and observes a significant improvement in agent performance. Experiment 3 reveals that the AI–human coach assemblage outperforms either the AI or human coach alone. This assemblage can harness the hard data skills of the AI coach and soft interpersonal skills of human managers, solving *both* problems faced by bottom- and top-ranked agents. These findings offer novel insights into AI coaches for researchers and managers alike.

Keywords: Artificial Intelligence, AI Coach, Salesforce Management, Sales Training, New Technology, Aversion, Information Overload

As the data-driven capability of AI improves (Brynjolfsson and Mitchell 2017), firms are exploiting AI coaches to train sales agents. AI coaches are computer software that leverage deep learning algorithms and cognitive speech analytics to analyze sales agents' conversations with customers and provide training feedback to improve their job skills. Due to their high computation power, scalability, and cost efficiencies, AI coaches are more capable of generating data-driven training feedback to agents than human managers. Indeed, MetLife, an insurance giant, adopted an AI coach named Cogito to offer training feedback to its call center frontline employees to improve customer service skills (Council 2019). Similarly, Zoom used its AI coach, Chorus, to offer on-the-job training to its sales force (Matheny 2019).

However, are there caveats in leveraging AI coaches for sales training? Precisely because of the big data analytics power of AI coaches, one concern is that feedback generated by the technology may be too comprehensive for agents to assimilate and learn, especially for bottom-ranked agents. Further, despite their superior "hard" data computation skills, AI coaches lack the "soft" interpersonal skills in communicating the feedback to agents (Shellenberger 2019), which is a key advantage of human managers (e.g., Jackson 1988; Daniels 2003). The lack of soft skills may result in an aversion to the AI coach (Dietvorst, Simmons, and Massey 2015, 2018; Srivastava 2019), which hampers their learning and performance improvement. Indeed, the design of AI coaches often focuses on information generation, but less on learning by agents who may differ in learning abilities (Roose 2019). Thus, it would be naïve to expect a simple, linear impact of AI coaches, relative to human managers, across heterogeneous sales agents.

Against this background, we address several research questions: (1) Which types of sales agents, bottom-, middle-, or top-ranked, benefit the most or the least from AI vis-à-vis human coaches? Is the incremental impact of AI coaches on agent performance heterogeneous in a non-

linear manner?; (2) What is the underlying mechanism? Does learning from the training feedback account for the impact of AI coaches?; and (3) Can an assemblage of AI and human coach qualities to circumvent the caveats and improve the sales performance of distinct types of agents?

To answer these questions, we conducted a series of randomized field experiments with two fintech companies. In experiment 1, a total of 429 agents were randomly assigned to undergo on-the-job sales training with an AI or human coach. Results show that the incremental impact of the AI coach over human coach is heterogeneous in an inverted-U shape. While middle-ranked agents improve their performance by the largest amount, both bottom- and top-ranked agents show limited incremental gains. Results suggest that this pattern is driven by a learning-based underlying mechanism: bottom-ranked agents encounter the most severe information overload problem with the AI coach, which leads to less learning from the coaching feedback and thus limited gains. By contrast, top-ranked agents display the strongest AI aversion problem, which obstructs their incremental learning and performance.

The slim improvement in bottom-ranked agents is an obstacle for the AI coach adoption because they have the largest room and most acute needs to sharpen their job skills. Thus, we re-designed the AI coach by restricting the amount of feedback provided to bottom-ranked agents. With a separate sample of 100 bottom-ranked agents, experiment 2 affirmed a substantial improvement in agent performance when the AI coach is restricted. Furthermore, experiment 3 tackles the limitations of either AI or human coaches alone by examining an AI–human coach assemblage, wherein human managers communicate the feedback generated by the AI coach to the agents. A new sample of 451 bottom- and top-ranked agents were randomly assigned to the AI coach, human coach, and AI–human coach assemblage conditions. The results suggest that

both bottom- and top-ranked agents in the AI–human coach assemblage condition enjoy higher performance than their counterparts in the AI coach alone or the human coach alone condition. In addition, bottom-ranked agents gain more performance improvement than top-ranked agents with the hybrid of AI and human coaching. Thus, this assemblage harnessing the soft communication skills of human managers and hard data analytics power of AI coaches can effectively solve both problems faced by bottom- and top-ranked agents.

Our research makes three key contributions to the literature. First, it is among the first to uncover the nuanced value of AI for salesforce management: the AI coach can be deployed to assist agents to learn and improve performance, rather than displace them. Our work extends prior literature on the negative impact of AI (Frey and Osborne 2017; Acemoglu and Restrepo 2017; Bessen et al. 2019) and customers' aversion to AI automation and algorithm (Dietvorst, Simmons, and Massey 2015; Mende et al. 2019; Luo, et al. 2019). Second, our results on the heterogeneous inverted U-shaped effects of the AI coach and the learning mechanism are important because they identify the distinct challenges faced by bottom- and top-ranked agents when trained by AI vis-à-vis human managers. These results refute a linear view of the effectiveness of deploying AI coaches in salesforce management. Third, we highlight a novel AI–human coach assemblage that outperforms either the AI or human coach alone condition. Adopting AI for agent training should also avoid the single-minded view of relying on AI coaches solely, or replacing human coaches with the autonomous data-driven machines completely. Rather, designing an assemblage where the smart machines assist human managers proves to be most effective in training salespeople for optimal performance.

Managerially speaking, our research empowers companies to tackle the challenges they may encounter when investing in AI coaches to train distinct types of agents. We show that,

instead of simply applying the AI coach to the workforce, managers ought to prudently design it for targeted agents. Moreover, companies should be aware that AI and human coaches are not dichotomous choices. Instead, an assemblage between AI and human coaches engenders higher workforce productivity, thus galvanizing companies to reap substantially more value from their AI investments.

AI Versus Human Coaches

A core advantage of AI technologies is their “hard” data computation skills. AI’s distinctive strength lies in processing “big data” and learning the latent patterns hidden in the structured and unstructured data (Davenport and Ronanki 2018; Luo et al. 2019; Puntoni et al. 2020). The backbone AI technologies consists of deep learning-based technologies such as Natural Language Understanding (NLU), Automatic Speech Recognition (ASR), Text-to-Speech Synthesis (TTS), Voice-Operated Switch (VOS), and Media Resource Control Protocol (MRCP). AI has been recognized as most suitable for tasks that require heavy processing of text, speech, image, and video data (Brynjolfsson and Mitchell 2017; Sundblad 2018). As AI technologies become increasingly sophisticated, they are able to perform many tasks conventionally carried out by humans, complement humans, and even outperform humans (McKinsey Global Institute 2018). For example, in the context of outbound sales calls, AI can understand customers’ queries and serve them in natural language conversations more competently than inexperienced workers do (Luo et al. 2019). Further, in e-commerce settings, AI can effectively handle data-intensive tasks such as machine translation and product recommendations (Brynjolfsson, Hui, and Liu 2019; Sun et al. 2020).

In the context of on-the-job sales training, a coach’s task is to review agents’ past conversations with customers and then provide feedback that enables them to learn sales skills

and improve future performance (Weitz, Sujan, and Sujan 1986; Román, Ruiz, and Munuera 2002). This task can be highly data intensive because coaches need to (1) listen to the speech data to identify mistakes in agents' conversations in serving customers and (2) provide specific solutions to rectify each of the mistakes. Because AI coaches can process high levels of speech data more effectively, they can detect a broader range of mistakes in conversations than human managers. Further, AI coaches are trained with a vast amount of past call data tagged as best and worst practices to persuade customers, so they can provide more solutions for each mistake identified in speech data processing. Overall, AI coaches' hard data computation skills suggest their relative advantages over human managers in generating feedback for sales agents.

In contrast, human managers' distinctive strength lies in their "soft" interpersonal communication skills. Specifically, interpersonal communication skills (e.g., interpersonal empathy, encouragement, adaptivity, and acknowledgement) are at the heart of human advantages over machines and are where AI falls short (Brynjolfsson and Mitchell 2017; Deloitte 2017; Deming 2017; Davenport and Ronanki 2018). Effectively conveying feedback to agents is pivotal for them to learn from the coaching information to improve job performance (Sujan, Weitz, and Kumar 1994; Román, Ruiz, and Munuera 2002). Successful communications often hinge on the degree to which coaches can adapt feedback to agents' learning capability and offer interpersonal support such as empathy, acknowledgement, and encouragement (Simon 1955; Tversky and Kahneman 1974). Such interpersonal skills of human coaches can reduce agents' resistance to coaching feedback (gain more trust) and overcome their learning barriers (Román, Ruiz, and Munuera 2002; Atefi et al. 2018). Thus, human coaches' distinctive interpersonal skills indicate their relative advantages over AI coaches in communicating feedback to agents.

The Inverted U-Shaped Effects of AI Coaches on Sales Agents

We posit that the incremental impact of AI coaches over human coaches is heterogeneous across agents in an inverted-U shape: while middle-ranked agents learn and improve their performance by the largest amount, both bottom- and top-ranked agents show limited incremental gains.

To begin with, bottom-ranked agents may encounter the most severe information overload problem associated with AI coaches (versus human coaches). The information processing theory has long viewed employees as information processors with limited capability in the literature (e.g., Simon 1955; Newell and Simon 1972; Tversky and Kahneman 1974; Fiske and Taylor 1991). There is also mounting evidence that too much information may introduce an information overload problem, which results in poor learning and performance (Jacoby 1974; Scammon 1977). In this sense, because an AI coach has advantages in data analytics and computational power compared with a human coach, its more comprehensive feedback may convey too much information for agents to digest. However, this information overload problem of AI versus human coaches is expected to be most severe among *bottom-ranked agents*. This is because bottom-ranked agents, who are inexperienced and the least skillful, tend to make many mistakes when serving customers. The data-driven AI coaches designed to spot mistakes and provide solutions are therefore likely to generate larger amounts of feedback (in terms of both breadth and depth) for bottom-ranked agents relative to human coaches. However, because of their lack of experience and skills, bottom-ranked agents are most likely to be overwhelmed by the comprehensive feedback from AI coaches and have the hardest time digesting such feedback, which then hampers their learning from coaching feedback and subsequent performance improvement.

Further, top-ranked agents may display the strongest aversion to AI versus human coaches. According to the information processing theory, the lack of soft interpersonal communication skills is a major roadblock to individual learning and performance improvement (Newell and Simon 1972; Sujan, Weitz, and Kumar 1994). Compared with human managers, the AI coach lacks interpersonal skills, which may result in more aversion to its coaching feedback (Dietvorst, Simmons, and Massey 2018; Srivastava 2019). People tend to trust humans and resist AI, even if the former does not perform as well as the latter, since machines are perceived to be cold and less empathetic without interpersonal communication skills (Önkal et al 2009; Eastwood, Snook, and Luther 2012; Putoni et al. 2020). This AI aversion has negative effects in domains such as consumer experience, hiring, learning, prediction, and medical diagnoses (Chaiken 1980; Ratneshwar and Chaiken 1991; Edmondson, Kramer, and Cook 2004; Puntoni et al. 2020).

In line with this literature, the aversion to the AI versus human coach among agents may adversely impact their learning and performance. However, such aversion to the AI versus human coach is expected to be the strongest for *top-ranked agents*. This is because they are skillful already and have their own views of the best practices for sales tasks. Indeed, Dietvorst, Simmons, and Massey (2015) point out that competent workers will have a stronger aversion to AI algorithms. Echoing this, Logg, Minson, and Moore (2019) find that experts trust advice from machines less than humans. Moreover, it has been found that experienced employees have more desire for autonomy and control (Denton and Kleiman 2001), which induces more aversion to machines (Burton, Stein, and Jensen 2020). Thus, top-ranked agents would have the strongest aversion to AI relative to human coaches, thus obstructing their learning and performance gains.

In summary, bottom-ranked agents encounter the most severe information overload problem with AI relative to human coaches, and top-ranked agents display the strongest problem of aversion to AI versus human coaches, although all agents may suffer from these two problems to some degree. Indeed, *middle-ranked agents* are more experienced than bottom-ranked agents in assimilating the training information from AI coaches (thus have a less severe information overload problem), and are not too skillful as their top-ranked peers to resist the useful personalized training feedback from AI coaches (thus have a less severe AI aversion problem). Hence, we expect that the incremental learning from the AI coach and subsequent performance improvement to be the strongest for middle-ranked agents compared with bottom- and top-ranked agents, in an inverted-U shape.

H1: The incremental impact of the AI coach over human coaches is in an inverted-U shape: while middle-ranked agents improve their sales performance by the largest amount, both bottom- and top-ranked agents show limited gains.

H2: Agents' learning from coaching feedback mediates the inverted U-shaped relationship in **H1**.

Figure 1 provides an overview of our three field experiments to test these and other predictions. The first experiment tests **H1** and **H2**, which highlight the heterogeneous inverted U-shaped effects of AI coaches on agent performance, the learning-based underlying mechanism, and the problems faced by bottom- and top-ranked agents.

=====Insert Figure 1 about here=====

Field Experiment 1

Empirical Setting

To test these hypotheses, we conducted a randomized field experiment with a large fintech company in Asia. The company is specialized in providing financial services to

individual customers. It has around 3,500 employees and over 19 million customers with an annual sales revenue of over \$3 billion. The company has a mobile app-based platform, which offers personal loans to individual customers. Its customers are between 22 and 55 years old. Each customer may select a loan amount that ranges from \$200 to \$8000 and choose the number of months to pay back the loan (between 6 and 24 months). To apply for a loan, a customer needs to upload his/her ID, together with documents describing his/her financial information such as income, assets, existing loans, and credit score. Once the loan is approved, the customer can receive the money on the same day. The monthly interest rate ranges from 1% to 3%, depending on the loan amount, length of time to pay back the loan, and the risk factor of the applicant calculated by the firm. The loans are often used to buy new smartphones, TVs, and computers.

The company hires sales agents to make promotional calls to its existing customers. The targeted customers usually have a credible repayment history of their current loan, so the company wants to offer them a special deal for renewing the loan under similar terms (e.g., loan amount, interest rate, and installments). Each customer receives at most one call from the company every three months. The tasks of sales agents include explaining the details of the promotion to the customer, answering questions from the customer about the financial product, and helping the customer prepare for the loan application. Every day, each agent is required to make about 50 effective sales calls to customers.¹ To ensure fairness, the customer calling lists are randomly assigned to the sales agents. The company records the number of successful loan applications and compensated agents accordingly.

¹ An effective sales call has to be answered by the customer and is not hung up immediately.

As successful sales promotion requires on-the-job training, the company hires managers to provide continuous sales training with call scripts and voice control techniques to sales agents before this experiment. Owing to the shortage and high marginal costs of hiring human managers to coach sales agents, the company adopts an AI coach. The prototype of the AI coach came from a high-tech platform with big data analytics skills and deep learning-based technologies such as Natural Language Understanding, Automatic Speech Recognition, Text-to-Speech Synthesis, Voice-Operated Switch, and Media Resource Control Protocol. This AI coach relies on a comprehensive best-practice knowledge bank to analyze agents' customer calls (unstructured audio data) and provide training feedback. The knowledge bank was created from a large volume of training data, namely, the company's historical audio recordings (in tens of terabytes) of sales calls.

The AI coach follows four steps to create and maintain the knowledge bank. First, it uses natural language processing techniques to convert the audio recordings into text scripts. The algorithm then conducts semantic parsing to convert the scripts into machine-understandable representations. In the third step, the AI coach applies deep learning models to identify context-dependent answers, not only those that maximize the predicted probability of the customers taking the promotion and renewing the loan (i.e., good answer bank), but also those that are either ineffective or even reduced the loan renewal probability (i.e., bad answer bank). Further, the AI coach is engaged in dynamic learning such that it repeats the previous steps with new training data. Once trained, the AI coach can automatically identify mistakes in the sales conversations (e.g., ambiguous words, unprofessional responses, overpromising, etc.) and generate comprehensive and personalized feedback (suggestions to remedy each mistake) to each sales agent to improve his/her job skills to persuade customers. While the development of the

prototype of the AI coach takes years, it can be applied to specific company settings relatively quickly (about one to two weeks). The AI coach is also cheaper to maintain. Specifically, the average cost of a human coach was about 15,000-20,000 local currency (\$2,100-2,820 in USD) per month including the insurances, social security, and retirement benefits required by the law. By contrast, the total cost of maintaining the AI coach system is around 10,000 local currency (\$1,410 in USD) per month—only half of the cost of hiring a human manager.

Experiment Design

In the field experiment, the company randomly assigned 429 sales agents to receive on-the-job training feedback from two types of coaches: the AI or human coaches. Within each type, there were three kinds of sales agents: bottom-, middle-, and top-ranked agents based on their previous performance before the experiment. Thus, our experiment design had a total of six conditions. The bottom-ranked agents had a previous performance in the bottom 20th percentile (usually interns and rookies in the company), the top-ranked agents were in the top 20th percentile, and the rest were the middle-ranked agents. To assure a balanced sample, the company randomly selected around 70 sales agents for each of the six experiment conditions.

For each agent, the sales coach (human or AI) listened to the sales call audios and provided training feedback that would improve his/her skills to handle future calls to persuade customers. Although the AI coach can scale up to scan thousands of sales calls simultaneously, to assure a fair comparison, the AI and human coaches across the six experiment conditions listened to the same number of randomly sampled calls for each agent. In the human coach condition, ten human managers were randomly assigned to train the three types of sales agents to control for manager-specific effects (e.g., popularity among agents). According to the company, all human managers were seasoned experts in the fintech industry with extensive sales training

experience. Each human manager was in charge of 20-25 sales agents, and every day a manager was required to listen to 5 randomly selected sales calls for each of his/her assigned agents. Thus, each manager listened to about 100 to 125 calls every day and spent about 3-4 minutes on each sales call (which typically lasted for 1-3 minutes), a normal workload for sales coaches in the industry. The compensation plan of human managers was designed such that about 70-80% of their income was linked to the performance of the sales agents assigned to ensure that the human managers had financial incentives to provide high-quality training feedback to the sales agents (thus making our results on the incremental impact of AI coaches more conservative).

Further, to rule out more alternative explanations, both AI and human coaches provided feedback with the same frequency, timing, and format. Specifically, daily training feedback was provided to the agents via email at 9 am on the next working day. For both AI and human coaches, the feedback included the text excerpted from the agent's sales conversation, sales mistakes the agent made during the conversation, and suggested solutions to each mistake. Web Appendix 1 provides examples of the training feedback from the two types of coaches. Although all coaches could identify mistakes in the sales conversation and provide effective suggestions, the AI coach was able to detect more mistakes and summarize more solutions. This also supports the face validity of our study: as the training task here is data-driven (i.e., analyzing the unstructured data on sales call conversations between the agent and customers), the AI coach should generate more comprehensive training feedback by identifying more mistakes and generating more solutions to remedy each mistake compared with human coaches. Because the AI coach is a cutting-edge technology newly adopted by the company, no agents had any prior AI coach experience before our experiment. In our experiment, all agents underwent the same amount of time in training (one month) for both the AI and human coach groups.

Data

We collected data on sales agents' performance, demographics, and voice characteristics for all their sales calls during the experiment month. We also surveyed the sales agents to understand their perceptions of the training feedback they received from the AI or human coaches right after the experiment. Table 1 presents the definitions of these variables and summary statistics.

=====Insert Table 1 about here=====

The performance of a sales agent was measured by the average purchase rate (in %), or the daily ratio of the sales calls successfully converted into loan renewal to the total sales calls averaged over the experiment month. The average purchase rate of the sales agents in the experiment month was around 14%, with a slightly higher rate in the second half of the month. Also, the performance was 12% in the month prior to the experiment.

In our study, the sales agents were young with an average age of 22 years old. About one-third of the agents were male. They were relatively well educated, with about 60% possessing a college degree. The industry had a high churn rate, as the majority of the sales agents in the sample had a tenure of less than 3 months. The general awareness of AI was low, as only 5% of the agents had prior experience of using AI-powered personal assistants (e.g., Tmall Genie offered by Tmall.com and Xiao AI offered by the company Xiaomi) before the experiment.

The data on the voice metrics was obtained from the AI algorithm. It is important to note that, although the agents from the human coach group did not receive feedback from the AI coach, their sales conversations were still monitored by the AI algorithm to provide consistent voice metrics across the groups. The agents were relatively professional in their sales conversations, as mistakes were only identified in around 6% of their sales calls, and about 94%

of their calls had an overall positive emotion. In addition, customers displayed a positive emotion in 43% of the calls and used objection/refuse words in 20% of the times in the conversations.

Further, we conducted a randomization check on the pre-experiment performance and demographic characteristics of the sales agents. Results in Web Appendix 2 suggest no statistically significant difference between the AI coach and human coach conditions across all the bottom-, middle-, and top-ranked agent levels. Thus, the randomization check was satisfied.

Results

Model-free evidence. We compared the performance of the sales agents in the AI versus human coach conditions to assess the incremental impact of the AI coach across the bottom-, middle-, and top-ranked agents. Figure 2 presents the model-free evidence. The results in the histograms suggest an inverted U-shaped impact of the AI relative to human coach: middle-ranked agents improved their performance by the largest amount (because the distribution of the performance of agents in the AI coach condition is on the right, and there is no overlap with that of agents under the human coaches). In addition, bottom-ranked agents showed somewhat limited performance gains (since there is some overlap between the two distributions), while top-ranked agents showed even less gains (because the two distributions are highly overlapped).

Further, the Panel A bar graph of Figure 3 shows that all agents attain higher performance under the AI relative to human coach ($p < .01$). In addition, the Panel B line graph of the same figure clearly supports that the incremental impact of the AI over human coach is in an inverted-U shape.

=====Insert Figure 2 and Figure 3 about here=====

Regression analyses results. To further control for the agents' individual characteristics and to quantify the relative impact of the AI versus human coaches across different agent types, we conducted a regression analysis.

$$\begin{aligned}
 \text{Purchase Rate}_i &= \alpha_0 + \alpha_1 \text{AICoach}_i + \alpha_2 \text{AICoach}_i \times \text{Bottom}_i \\
 &+ \alpha_3 \text{AICoach}_i \times \text{Top}_i + \alpha_4 \text{Bottom}_i + \alpha_5 \text{Top}_i + \alpha_6 \text{Age}_i \\
 &+ \alpha_7 \text{Male}_i + \alpha_8 \text{College}_i + \alpha_9 \text{Tenure2}_i + \alpha_{10} \text{Tenure3}_i \\
 &+ \alpha_{11} \text{AIExperience}_i + \epsilon_i,
 \end{aligned} \tag{1}$$

where Purchase Rate_i is the sales performance of the i th sales agent, measured by the daily ratio of the sales calls successfully converted into loan renewal to the total sales calls (in % and averaged over the experiment month). AICoach_i is the indicator variable for the random assignment of the AI coach (=1, and 0 for the human coach). Bottom_i is the indicator variable for the bottom-ranked agents whose previous performance was in the bottom 20th percentile (usually interns and rookies in the company) before the experiment. Top_i is the indicator variable for top-ranked agents whose previous performance was in the top 20th percentile. The rest of the agents were middle-ranked agents. In this model, the performance of middle-ranked agents from the human coach group is the baseline. We further interacted AICoach_i with Bottom_i and Top_i to allow for the heterogeneous *incremental* effect of the AI coach. Therefore, these two interaction terms capture the extent to which the causal incremental effect of the AI over human coaches for bottom- and top-ranked agents was different from that for their middle-ranked peers. Age_i represents the age of the agent. Male_i captures the fixed effect of male sales agents. College_i denotes a college education. Tenure2 and Tenure3 correspond to a tenure of 4-6 months and more than 6 months, respectively (tenure less than 3 months is the baseline). AIExperience_i equals one if the agent had prior experience of using AI-powered personal

assistants (e.g., Tmall Genie offered by Tmall.com, Xiao AI offered by the company Xiaomi), not the AI coach, before the experiment. ϵ_i is the error term.

Table 2 reports the results across the sales performance from different time periods (the entire month, the first half, and the second half of the experiment month). The coefficient of $AICoach_i$ is consistently positive and significant, indicating a positive incremental effect of the AI over human coaches on middle-ranked agents. This finding is reasonable because relative to human managers, AI coaches have advantages in hard data computation skills and thus generate more comprehensive training feedback to benefit sales agents. The coefficient of $AICoach_i \times Bottom_i$ is negative and significant, suggesting that bottom-ranked agents gain less from the AI coach than middle-ranked agents. Interestingly, the coefficient of $AICoach_i \times Top_i$ is more negative than that of $AICoach_i \times Bottom_i$ ($p < .01$), suggesting that the top-ranked agents gain the least from the AI coach.² Thus, these results confirm Figure 3's model-free evidence for the inverted U-shaped impact of the AI over human coach and strongly support **H1**: while middle-ranked agents improve their performance by the largest amount, both bottom- and top-ranked agents show limited gains.³

=====Insert Table 2 about here=====

Learning mechanism. To measure learning from coaching feedback with objective behavior data, we used the AI algorithm to analyze the sales conversations audio data on all agents (from both treatment and control groups) during the experiment and created a variable named $Mistakes_i$. This variable captures an agent's average proportion of sales calls with

² As a robustness check, we also estimated Differences-in-difference (DID) models by utilizing the time-series variation in sales agents' performance in the months before and during the experiment. The DID results are consistent (see Web Appendix 3).

³ We observed insignificant moderating effects of other variables (see Web Appendix 4).

mistakes identified by the AI coach during the experiment month (a value of 10% means that 10% of the agents' sales calls during the experiment had sales mistakes as identified by the AI algorithm). This variable is a proxy for agents' learning from the training feedback during the experiment period: the more agents learned from coaching feedback, the fewer mistakes they would make in their sales calls. To directly test the significance of the moderated mediation effects, we leverage the PROCESS Model 8 with 5,000 bootstrap replications (Hayes 2013). As summarized in Figure 4, the results suggest that under the AI versus human coach, both bottom- and top-ranked agents made significantly more sales mistakes than middle-ranked agents did ($p < .01$). Also, sales mistakes have a negative and significant relationship with purchase rates, as expected. Thus, the moderated effect of the AI versus human coach on the performance of the agents of different types is significantly mediated by the learning proxy of sales mistakes.⁴

Together, these results strongly support our **H2**.

=====Insert Figure 4 about here=====

We further explore the time-series variation in agent performance to offer evidence for the learning mechanism. Specifically, if the learning mechanism in H₂ is true, then we should observe different learning curves to reflect the performance gains throughout the course of the experiment. That is, the performance impact of the AI coach on agents should happen not immediately, but rather gradually in a learning curve, and each type of agents may have different shaped curves. As shown in Figure 5, across all agent types, we observe an incremental improvement under the AI relative to human coach (all the learning curves go upward over time). Such improvement was slightly larger in the first half of the month than that in the second

⁴ The indirect effects of *AI Coach* are 3.20, 6.10 and 0.60 for bottom-, middle- and top-ranked agents, respectively. Using middle-ranked agents as the baseline, the effect sizes of moderated mediation are -2.91 and -5.51 for bottom- and top-ranked agents, respectively. All estimates are significant at 0.05 level. More details of the moderated mediation analysis can be found in Web Appendix 5.

half, which is in line with a typical learning curve. More importantly, we observe the greatest learning progress and performance gains among the middle-ranked agents (dashed line), followed by bottom- and top-ranked agents (solid and dotted lines, respectively). Thus, these learning curve results corroborate **H₁** and **H₂**.

=====Insert Figure 5 about here=====

AI coach effects on customer satisfaction. We also checked the effects of the sales coach on customer satisfaction as proxied by customer sentiment (the percentage of the overall positive customer sentiment as detected by the AI algorithm) and customer objections (the percentage of the customer's words of objection such as no, don't, won't, etc.). Columns 3-5 in Table 3 show that the AI coach helped improve the sentiment in the sales conversation for customers. Further, we observe a similar inverted U-shaped effect: the AI coach promoted positive customer sentiments and reduced customer objections the most for middle-ranked agents, rather than the bottom- and top-ranked agents, thus adding more evidence for **H₁** and **H₂**.⁵

Ruling out the alternative explanation of working hard. Prior literature (Weitz, Sujan, and Sujan1986) suggests that salespeople can improve their performance by either working smarter (enabled by learning) or working harder (simply by making more sales calls). Thus, we checked the working hard behavior with two outcomes: *CallingTime_i* (an agent's average daily active sales calling minutes) and *NumberofCalls_i* (the average daily attempted sales calls). As shown in Columns 1 and 2 of Table 3, the coefficients of AI Coach and the interaction terms are insignificant ($p > .10$), thus ruling out the alternative explanation of working hard.⁶

⁵ We also tested agent performance conditional on the matched feedback volume and found consistent results (see Web Appendix 6).

⁶ We also considered the alternative explanations of the Hawthorne effect and "water cooler" effect. As shown in Web Appendix 7, these effects are unlikely to threaten our results.

=====Insert Table 3 about here=====

Evidence for the problems of information overload and aversion to the AI coach. To directly show the evidence for the problems of information overload and aversion to the AI coach, we leveraged survey data on the agents' perceptions of the feedback from the AI vis-à-vis human coaches. Results in Columns 6 and 7 in Table 3 (see also the top 2 charts in Figure 6) confirm that on average, the agents perceived the feedback from the AI coach to be more comprehensive than that from human coaches in terms of both breadth (i.e., number of mistakes identified) and depth (i.e., the number of suggested solutions to remedy each mistake) ($p < .01$). Yet, the more comprehensive feedback generated by the AI versus human coach does not mean that all agents would learn from coaching feedback and increase their performance equally. As shown in Column 9 in Table 3 (see also the bottom right chart in in Figure 6), bottom-ranked agents felt most overloaded by the comprehensive feedback from the AI versus human coach. In other words, bottom-ranked agents indeed faced the most severe information overload problem to learn and benefit from the more comprehensive feedback by the AI versus human coach (accounting for the left side of the inverted-U in H_1). By contrast, as shown in Column 8 in Table 3 (see also the bottom left chart in Figure 6), top-ranked agents felt the highest aversion to the AI versus human coach and thus had limited performance gains (accounting for the right side of the inverted-U in H_1). Indeed, middle-ranked agents may learn and improve the most from the comprehensive feedback by the AI versus human coach, as they suffered less from either the information overload problem (relative to bottom-ranked agents) or the problem of aversion to the AI versus human coach (relative to top-ranked agents). Thus, these findings provide more empirical support for **H₁** and **H₂**.

=====Insert Figure 6 about here=====

Discussion

Experiment 1 found that although middle-ranked agents experienced the greatest improvement in their performance, both bottom- and top-ranked agents attained limited gains from the AI versus human coach. This inverted-U pattern was driven by a learning-based mechanism: bottom-ranked agents encountered the most severe information overload problem with the AI versus human coach, while top-ranked agents had the strongest aversion to the AI relative to human coach. To solve these problems, we redesigned the AI coach by restricting it in experiment 2 and by proposing an AI–human coach assemblage in experiment 3.

Field Experiment 2: Restricting AI Coach Feedback

The limited performance gain for bottom-ranked agents is clearly an obstacle for adopting the AI coach. This is because these agents have the largest room to improve their sales skills and the most acute need for on-the-job training. The survey data in experiment 1 also support that the information overload problem with the AI versus human coach was indeed a key challenge that hampered the learning and performance of bottom-ranked agents. If it was the information overload problem of the AI versus human coach that obstructed bottom-ranked agents' learning and performance (Weitz, Sujan, and Sujan 1986; Sujan, Weitz, and Kumar 1994; Slater and Narver 1995), then we can tackle this problem by restricting the amount of feedback from the AI coach to bottom-ranked agents. That is, the information overload problem faced by bottom-ranked agents motivates us to engage in a redesign of AI coaches by restricting the amount of feedback. It is expected that relative to the unrestricted counterparts, the restricted AI coaches with less feedback amount will reduce the information overload problem and, through the reduction in information overload, have a positive impact on bottom-ranked agents' sales performance. This discussion leads to the following hypotheses:

H_{3a}: An AI coach that offers less (versus more) feedback has a positive impact on bottom-ranked agents' sales performance.

H_{3b}: The reduction in information overload mediates the relationship in **H_{3a}**.

Experiment Design

In this experiment, a new sample of 100 bottom-ranked sales agents were selected from the same company. These agents did not participate in experiment 1 before and were unaware of this follow-up field experiment until it began. We randomly assigned half of the selected sales agents into the control group, in which they received feedback from the AI coach similar to that in experiment 1 (unrestricted feedback from the AI coach). The other half of the agents were assigned into the treatment group, in which they received less feedback from the AI coach each day (restricted feedback from the AI coach). Specifically, for each sales mistake identified, the AI algorithm ranked all the potential feedback solutions according to its importance and reported the one that is predicted to have the largest impact on purchase rates. Experiment 2 lasted for another month, and all the data were collected similar to in experiment 1.

Data and Empirical Results

The summary statistics of the data are reported in Web Appendix 8. In this appendix, we carried out another randomization check and found the characteristics of the sales agents from the treatment group (AI coach with restricted feedback) and the control group (AI coach with unrestricted feedback) were not statistically different.

Next, we ran regression models to test the impact of reducing the amount of feedback from the AI coach. The empirical specification is as follows:

$$\begin{aligned} \text{Purchase Rate}_i &= \alpha_0 + \alpha_1 \text{AI Coach(Restricted)}_i + \alpha_2 \text{Age}_i + \alpha_3 \text{Male}_i \\ &+ \alpha_4 \text{College}_i + \alpha_5 \text{AI Experience}_i + \epsilon_i \end{aligned} \quad (2)$$

where $AICoach(Restricted)_i$, is an indicator variable that takes the value of 1 if the sales agent belongs to the AI coach group with restricted feedback group and 0 otherwise. As shown in Table 4, the coefficient of $AICoach(Restricted)_i$ is positive and significant ($p < .001$). Thus, for the bottom-ranked agents, receiving the restricted, rather than all, feedback from the AI coach was more effective for improving their job performance. On average, with the restricted AI coach, the purchase rate increased by approximately 4.6 percentage points, a 50% improvement in performance. Thus, these results provide causal evidence for **H_{3a}**.

=====Insert Table 4 about here=====

In addition, we tested the mediational role of the reduction in information overload via the PROCESS Model 4 with 5,000 bootstrap replications (Hayes 2013). Results suggest that the AI coach with less (versus more) feedback indeed helps reduce the information overload problem faced by bottom-ranked agents ($p < .01$). Also, information overload negatively affects agent performance ($p = .06$), as expected. The mediational effect of information overload is statistically significant at 10% level, thus marginally supporting **H_{3b}**.⁷

These findings with causal evidence also support the assumption of **H₁**: bottom-ranked agents indeed suffered from an information overload problem with the AI coach, which limited their sales performance gains. Nevertheless, a redesign of AI coaches by restricting the amount of feedback can effectively solve the information overload problem faced by bottom-ranked agents and improve their sales performance. However, this experiment did not tackle top-ranked agents' problem of strongest aversion to the AI versus human coach, which is addressed in experiment 3.

⁷ Details of the Mediation Analysis can be found in Web Appendix 9. The marginal significance may arise due to the small sample size ($n=100$) here.

Field Experiment 3: The AI–Human Coach Assemblage

In field experiments 1 and 2, either the AI or human coach was deployed, but using either coach alone has its own inherent disadvantages as discussed previously. Therefore, experiment 3 designed an assemblage in order to harness the advantages of both AI and human coaches. Specifically, we propose an AI–human coach assemblage wherein human managers communicate the feedback generated by the AI coach to sales agents. That is, the AI coach analyzes the speech data on sales conversations and generates the data-driven feedback, while the human manager communicates the feedback to agents. This assemblage can leverage the advantages of both types of coaches: the hard data skills of the AI coach (Brynjolfsson and Mitchell 2017; Wilson and Daugherty 2018; Luo et al. 2019) and soft interpersonal skills of human managers (Newell and Simon 1972; Sujan, Weitz, and Kumar 1994; Kannan and Bernoff 2019).

This AI–human coach assemblage helps solve multiple problems at the same time. On the one hand, because human managers are proficient in explaining the comprehensive feedback generated by AI with interpersonal communication skills (e.g., encouragement, empathy, and adaptivity), bottom-ranked agents will be less likely to face the information overload problem and thus can learn and improve more (Jacoby 1974; Weitz, Sujan, and Sujan 1986). In addition, because human managers can converse the feedback generated by AI with interpersonal communication skills (e.g., acknowledgement, empathy, and adaptivity), top-ranked agents should also display less resistance to coaching feedback and hence enjoy higher performance gains (Dietvorst, Simmons, and Massey 2015, 2018; Luo et al. 2019; Srivastava 2019). This discussion suggests that relative to either the AI or human coach alone condition, an AI–human coach assemblage has a positive impact on bottom- and top-ranked agents' performance. Further,

we expect bottom-ranked agents to benefit more from this assemblage than top-ranked agents. This is because bottom-ranked agents are the least skillful and have more room to learn and improve under the AI–human coach assemblage; by contrast, top-ranked agents can also improve but to a lesser degree because they are most skillful already (e.g., near the top limit). As a result, we offer the following hypotheses:

- H_{4a}:** Relative to either the AI or human coach alone, an AI–human coach assemblage has a positive impact on bottom-ranked agents’ sales performance.
- H_{4b}:** Relative to either the AI or human coach alone, an AI–human coach assemblage has a positive impact on top-ranked agents’ sales performance.
- H_{4c}:** The incremental positive performance impact of the AI–human coach assemblage is amplified for bottom-ranked agents relative to top-ranked agents.

Experiment Design

We conducted this field experiment in a different company. The company operates in the same fintech industry and specializes in peer-to-peer (P2P) loan collection. In this company, sales agents call delinquent individual borrowers to collect overdue loan payments. The personal loans range from \$200 to \$8000. Because this company is also from the fintech industry, the background information of the sales agents, AI coaches, and experiment procedures in experiment 3 were similar to those in experiments 1 and 2.

However, the AI coach in this experiment performs a different training task: coaching sales agents to improve their loan collection skills. This different setting helps us boost the generalizability of our findings. Moreover, to extend experiment 1 where human managers provided feedback to the sales agents via emails in one-way communications, we allow human managers to communicate interpersonally in two-way communications in experiment 3. In this way, human coaches could leverage their soft interpersonal communications skills (thereby making it harder for the AI coach to outperform human coaches in this setting). More

specifically, while the AI coach still communicated with the sales agents via emailed feedback, human coaches in experiment 3 held a one-to-one face-to-face meeting with each of the assigned agents.⁸ This format allows for the managers and agents to interact with each other in the job training. In the AI–human coach assemblage, the AI coach assisted the human managers by listening to the agents’ sales conversations and generating the training feedback (similar to the AI coach alone group), and the human managers held a one-to-one face-to-face meeting with each of the assigned agents (similar to the human coach alone group). In this way, the assemblage group had not only the hard data analytical advantages of AI coaching, but also the soft interpersonal communication skills of human coaching.

Sale agents (n=451) were randomized to undergo loan collection training in three conditions: the human coach alone group, the AI coach alone group, and the AI–human coach assemblage group. In this month-long field experiment, 224 of them were randomly selected from bottom-ranked agents, and the rest were randomly selected from the top-ranked agents.⁹ We further randomly assigned these agents to the three conditions. Thus, we had six experiment conditions, and each condition had a roughly balanced sample in the data.

Data and Empirical Results

We summarize the variables in Web Appendix 10. This appendix also reports the results for the randomization check. Conditional on the pre-treatment performance level, agents from the three treatment groups were not significantly different from each other in all the pre-treatment characteristics. The data therefore passes the randomization check.

⁸ In Experiment 3, the feedback involved from the AI coach was the total amount given and not the truncated level. Also, depending on the volume and the content of the feedback, such a one-to-one face-to-face meeting may last for up to 10 minutes. However, most of the meetings lasted for about 2-5 minutes in our study.

⁹ This company did not agree to include middle-ranked agents in experiment 3 because, based on the first experiment, such agents benefited the most from the AI coach already and did not suffer seriously from the two problems faced by bottom- and top-ranked agents.

To examine the relative impact of the AI–human coach assemblage, we estimated a regression analysis with the following specification:

$$\begin{aligned} \text{DailyCollectionAmount}_i = & \alpha_0 + \alpha_1 \text{AICoachAlone}_i + \\ & \alpha_2 \text{AIHumanCoachAssemblage}_i + \alpha_3 \text{AICoachAlone}_i \times \\ & \text{Top}_i + \alpha_4 \text{AIHumanCoachAssemblage}_i \times \text{Top}_i + \\ & \alpha_5 \text{Top}_i + \alpha_6 \text{Age}_i + \alpha_7 \text{CommunityCollege}_i + \\ & \alpha_8 \text{College}_i + \alpha_9 \text{Tenure2}_i + \alpha_{10} \text{Tenure3}_i + \\ & \alpha_{11} \text{Tenure4}_i + \alpha_{12} \text{AIExperience}_i + \epsilon_i, \end{aligned} \quad (3)$$

where *DailyCollectionAmount_i* measures the sales agent’s average daily collection of customers’ loan repayment (averaged over the experiment month). *AICoachAlone_i* and *AIHumanCoachAssemblage_i* are indicator variables for agents from the respective groups (compared with the human coach alone group). We further interacted these two variables with *Top_i* to allow for these two treatment effects to differ among top- and bottom-ranked agents. Thus, the performance of bottom-ranked agents in the human coach alone group is the baseline.

Table 5 reports the regression results. Columns 1, 2, and 3 suggest that the coefficients of *AIHumanCoachAssemblage_i* are positive and significant, and are statistically larger than the coefficients of the *AICoachAlone_i* across the whole data sample, bottom-ranked agents, and top-ranked agents, respectively ($p < .01$). These results suggest that, for both top- and bottom-ranked agents, the assemblage of the AI and human coaches is more effective than either coach alone. This pattern is shown in Figure 7 on the incremental impact of the AI–human coach assemblage and the AI coach alone (over the baseline of the human coach alone). Thus, **H4a** and **H4b** are both supported.¹⁰

¹⁰ For bottom-ranked agents, the positive and significant coefficient of *AICoachAlone_i* in Column 2 suggests that the AI coach alone still outperforms the human coach alone, thus extending experiment 1 to a new setting of loan collection training. Further, for top-ranked agents, the negative and significant coefficients of *AICoachAlone_i* in Column 3 and *AICoachAlone_i × Top_i* in Column 4 are reasonable because top-ranked agents may hold an even stronger aversion to the AI coach relative to human managers providing *two-way* communications feedback in experiment 3 (versus managers providing one-way communications feedback in experiment 1).

=====Insert Table 5 and Figure 7 about here=====

Further, results in Column 4 suggest that there is a negative and significant coefficient of $IHumanCoachAssemblage_i \times Top_i$. Thus, while improving, top-ranked agents had a significantly smaller performance improvement than bottom-ranked agents when coached by the AI–human coach assemblage (also pictorially shown in Figure 7). Thus, H_{4c} is strongly supported by our data.

Discussion

Despite growing interests among firms, effectively leveraging AI coaches in sales training remains poorly understood. We conducted three randomized field experiments in two fintech companies to quantify the heterogeneous effects of AI coaches on agents' performance and redesigned AI coaches to solve the challenges faced by bottom- and top-ranked agents. Experiment 1 found that while middle-ranked agents experienced the greatest improvement in their performance, both bottom- and top-ranked agents gained little from the AI versus human coach. This inverted-U pattern was driven by a learning-based mechanism where bottom-ranked agents encountered the most severe information overload problem with the AI versus human coach, and top-ranked agents had the strongest aversion to the AI relative to human coach. To alleviate the information overload experienced by the bottom-ranked agents, experiment 2 redesigned the AI coach by restricting the feedback amount to these agents and revealed a substantial improvement in their sales performance. Further, experiment 3 moved beyond either the AI or human coach alone, toward combining them into the AI–human coach assemblage. This hybrid between AI and human coaches outperformed either coach alone in terms of training effectiveness for both bottom- and top-ranked agents. These findings provide broad implications for research and practice.

Contributions to Research

Our research proffers several contributions to the literature. This study is among the first to examine the effectiveness of AI coaches in sales training and uncover evidence that AI can be designed to complement, rather than substitute, employees in customer services. This is critical because it changes our knowledge on how AI may reshape the business landscape of serving customers in the real world. Recent studies noted customers' AI aversion. On the basis of lab experiments, they show that humanoid robots and AI automation threaten customers' own identity (Mende et al. 2019), personal consumption experience (Leung, Paolacci, and Paolacci 2018; Castelo, Bos, and Lehmann 2019), and perceived uniqueness (Longoni, Bonezzi, and Morewedge 2019). Incidentally, on the employee side, lab studies also noted the negative views among workers towards computer algorithms (Dietvorst, Simmons, and Massey 2015, 2018; Li, Liu, and Liu 2016). Warning against the "dark side" of AI, Acemoglu and Restrepo (2017) and Frey and Osborne (2017) point out the concerns of AI displacing human jobs. Advancing this literature, we address a timely and important topic on AI coaches. On the basis of multiple field experiments in real-word contexts, we uncover causal evidence for the positive effects of AI in coaching salespeople to better serve customers. When AI is designed to complement employees on their tasks, the negative consequences of job loss and workers' resistance to AI are mitigated. Indeed, AI technology can scale up and assist a large number of sales agents simultaneously to improve their job performance, especially when there is a shortage of human managers for sales training and coaching jobs.

Further, we extend the literature by uncovering the heterogeneous inverted U-shaped effects of the AI coach across various types of sales agents. While benefiting middle-ranked agents the most, AI coaches faced distinct challenges in training bottom- and top-ranked agents.

We show that bottom-ranked agents encountered the most severe information overload problem, whereas top-ranked agents had the strongest aversion to the AI coach. These findings are non-trivial because they rebut the naïve, linear view of the impact of AI in salesforce management. They also provide new insights into assessing the effectiveness of sales training by AI coaches vis-à-vis human managers. Earlier studies have focused on the average linear effect of sales training programs by human coaches (Román, Ruiz, and Munuera 2002; Atefi et al. 2018). Examining the heterogeneous effects of sales training by AI helps pinpoint the employee-specific challenges and then craft effective solutions. Researchers should not neglect the different training needs among sales agents and ought to pay high attention to the limitations of AI coaches. While bottom-ranked agents have the highest need for sales training, data-driven and computationally powerful AI coaches that provision too much information may not suit them. The more powerful the big data analytics skills of AI and the more comprehensive the feedback from AI, the more dysfunctional it might be for bottom-ranked agents. The data-driven advantage of AI might thus turn into a disadvantage when using AI to coach bottom-ranked agents (i.e., the good intention of adopting new AI technologies may end up with a bad outcome).

In addition, we contribute to the literature by revealing the learning-based mechanism for the performance impact of AI coaches. Prior research noted the importance of individual learning (working smarter; Weitz, Sujan, and Sujan 1986; Sujan, Weitz, and Kumar 1994; Slater and Narver 1995) for sales performance. We add to this literature by supporting the mediating role of agent learning in the heterogeneous impact of AI coaches on sales agent performance. Such findings are crucial because they show that the AI coach offering comprehensive training information does not automatically improve agent performance, unless it can foster their learning

as an intermediary outcome. Indeed, because bottom-ranked agents face barriers to assimilate the comprehensive feedback, they may learn more from the AI coach with restricted information provision (which eases the information overload problem) and thus boost their performance more robustly.

Moreover, we advance the literature by conceptualizing the AI–human coach assemblage and by revealing its superior impact relative to either the AI or human coach alone. Prior research notes the economic value of AI *alone* (i.e., fully autonomous without human involvement) in machine translation, conversational commerce, and outbound sales calls (e.g., Brynjolfsson, Hui, and Liu 2019; Luo et al. 2019; Sun et al. 2020). Extending this literature, we put forth the novel idea of the AI–human assemblage. We show that the assemblage where human managers communicate the feedback generated by the AI coach cannot only solve the acute problem of the information overload for bottom-ranked agents, but also turn AI aversion into appreciation for top-ranked agents. Thus, adopting AI should also avoid the single-minded view of relying on AI coaches solely (i.e., replacing human managers with the data-driven machines). Rather, an assemblage of both, where the smart machines assist rather than displace human managers, is the most effective tool for training salespeople for optimal performance. It is fruitful to explore new opportunities of the AI–human coach assemblage, as the AI and human coach each have distinct strength and weakness. Combining both can tether the benefits of AI’s hard data analytical abilities and human managers’ soft interpersonal communications skills. The economic value of AI coaches is substantially greater if the same AI technology is designed to assist human managers in the assemblage.

Managerial Implications

Our findings provide several useful insights to managers. First, sales training has long been regarded as an essential yet costly and challenging task for human managers (Chung, Park, and Kim 2019). Each year, companies lose over \$75 billion in revenues because of poor customer services (Teich 2019). More than \$1 trillion dollars have been invested in companies' call centers to handle over 265 billion customer calls annually and train agents to improve the quality of customer services. In 2016, the average spending on training per salesperson reached \$1,459, almost 20% more than that for other workers. Paradoxically, such high investment in sales training does not always translate into performance. About 80% to 90% of the investment is either ineffective or difficult to quantify (Schultz 2011). Such ineffective training is conventionally handled by human managers (Martin and Collins 1991; Christiansen et al. 1996; Dubinsky 1996; Román, Ruiz, and Munuera 2002). Fortunately, with the rise of AI, companies can leverage AI coaches to train agents and improve their sales skills and performance more effectively and efficiently. Unlike human managers who may suffer from physical fatigue and emotional fluctuations, the AI coach will not have bad days or toxic emotions in the repetitive sales job training, so it can handle the training tasks in a more consistent, predictable, and accurate manner (Pew Research 2018). In addition, AI coaches can solve another thorny problem in the industry—the limited supply of human managers to train inexperienced frontline employees. For the data-driven and recurring tasks of on-the-job sales training, AI can scale up quickly to train thousands of agents simultaneously with minimal marginal costs.

However, AI coaches are not a magic bullet; they still have limitations. While AI coaches can provide comprehensive feedback to individual agents, they do not automatically take into account the different challenges faced by sales agents (e.g., information overload and AI aversion). Bottom-ranked agents have the most acute needs for job skill training, yet they may

not learn from the information generated by AI coaches because of information overload. By contrast, top-ranked agents may have a strong aversion to AI coaches, which can also limit their learning and performance gains. Hence, instead of simply applying AI coaches to the workforce and waiting for the software to get smarter, firms should be aware of the limitations of AI coaches in meeting the distinct training needs of these heterogeneous agents. Firms can address these limitations and provide effective solutions to different types of agents by deploying targeted AI coaches for sales agents. For example, for bottom-ranked agents, firms can restrict information provisioning to deal with the information overload problem. For both bottom- and top-ranked agents, firms can combine AI coaches and human managers to train their salesforce more effectively.

Optimally, our result suggest that companies should adopt the assemblage of AI coaches and human managers. AI and human coaches are not substitutes, but rather complements to each other. In this assemblage, managers can focus on communications that are interpersonal, nuanced, and difficult to automate, while AI provides hard data computation skills and personalized feedback at a scale that improve managers' communications with their subordinates. Put it differently, in this modern era of AI, the classical interpersonal competencies of human managers remain crucially important. Overall, the AI-human coach assemblage allows firms to achieve a three-win scenario, where: (1) sales agents can attain greater learning and income; (2) managers can be freed from mundane and repetitive training tasks and spend more resources on tasks that require creativity, judgment, and leadership; and (3) companies can enjoy higher sales revenues.

Limitations and Future Research

There are several limitations in this study that suggest opportunities for future research. First, our empirical evidence on AI coaches is based on two tasks (loan promotion and collection) in a specific industry (fintech). It would be interesting to examine whether AI coaches will have similar effects in other settings. For example, do our findings still hold if sales agents promote other products such as durable goods in B2B settings? Can the AI–human coach assemblage be more or less productive in training sales agents to persuade business clients and key corporate accounts? Would AI coaches be as effective as human managers if agents conduct in-person businesses instead of making sales calls over the phone? Future research is called for to explore the generalizability and boundary conditions of our findings. Further, the effectiveness of AI sales coaches in other training formats is worth studying. For example, can the AI coach be more or less effective due to social influence and public exposure if its feedback to an agent is also observable to her colleagues? This practice may facilitate cross-learning among the employees with different personality, but at the same time increase the public exposure of their personal failures, which may lead to both positive and negative ramifications. Since our experiments lasted for just one month, our findings largely explain the short-run effects of the AI coach. It is thus important to explore the long-run effects. For example, with the help of AI coaches, how quickly can bottom-ranked agents become the top performers? In addition, as the dispersion in sales agents' performance shrinks with the adoption of AI coaches, how would managers adjust their salesforce recruitment and promotion strategies in the long run?

In conclusion, our research is an initial step in examining the caveats and solutions in the context of leveraging AI coaches for sales training effectiveness. We hope it can stimulate more work on this pivotal interface between AI technologies and sales agent performance.

References

- Acemoglu, Daron, and Pascual Restrepo (2017), “Robots and Jobs: Evidence from Us Labor Markets,” Working Paper No. 23285, National Bureau of Economic Research.
- Anderson, Janna, and Lee Rainie (2018), “Artificial Intelligence and the Future of Humans,” (accessed August 9, 2020), <https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans>.
- Atefi, Yashar, Michael Ahearne, James G Maxham III, D Todd Donavan, and Brad D Carlson (2018), “Does Selective Salesforce Training Work?” *Journal of Marketing Research*, 55 (5), 722–737.
- Bessen, James E., Maarten Goos, Anna Salomons, and Wiljan Van den Berge (2019), “Automatic Reaction-What Happens to Workers at Firms that Automate?” Working paper.
- Brynjolfsson, Erik, Xiang Hui, and Meng Liu (2019), “Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform,” *Management Science*, 65 (12), 5449-5460.
- Brynjolfsson, Erik, and Tom Mitchell (2017), “What Can Machine Learning Do? Workforce Implications,” *Science*, 358 (6370), 1530-1534.
- Burton, Jason W., Mari-Klara Stein, and Tina Blegind Jensen (2020), “A Systematic Review of Algorithm Aversion in Augmented Decision Making,” *Journal of Behavioral Decision Making*, 33 (2), 220-239.
- Castelo, Noah, Maarten W. Bos, and Donald R. Lehmann (2019), “Task-Dependent Algorithm Aversion,” *Journal of Marketing Research*, 56 (5), 809-825.
- Chaiken, Shelly (1980), “Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion,” *Journal of personality and social psychology*, 39 (5), 752.
- Chung, Doug J., Byoung G. Park, and Byungyeon Kim (2019), “The Comprehensive Effects of Salesforce Management: A Dynamic Structural Analysis of Selection, Compensation, and Training,” Working paper.
- Christiansen, Tim, Kenneth R. Evans, John L. Schlacter, and William G. Wolfe (1996), “Training Differences Between Services and Goods Firms: Impact on Performance, Satisfaction, and Commitment,” *Journal of Professional Services Marketing*, 15 (1), 47-70.
- Council, Jared (2019), “MetLife Says AI Is Improving Its Call Centers,” (June 3), *The Wall Street Journal*, WSJ PRO (June 3).
- Daniels, Sharon (2003), “Employee Training: A Strategic Approach to Better Return on Investment,” *Journal of Business Strategy*, 24 (5), 39-42.
- Davenport, Thomas H., and Rajeev Ronanki (2018), “Artificial Intelligence for the Real World,” *Harvard business review*, 96 (1), 108-116.
- Deloitte (2017), “Soft Skills for Business Success” (accessed August 09, 2020), <https://www2.deloitte.com/au/en/pages/economics/articles/soft-skills-business-success.html>.
- Deming, David J. (2017), “The Growing Importance of Social Skills in the Labor Market,” *Quarterly Journal of Economics*, 132 (4), 1593-640.
- Denton, David W., and Lawrence S. Kleiman (2011), “Job Tenure as a Moderator of the Relationship Between Autonomy and Satisfaction,” *Applied HRM research*, 6 (2), 105-114.
- Dietvorst, Berkeley J. (2015), “Algorithm Aversion,” University of Pennsylvania, PhD Dissertation.
- Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2015), “Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err,” *Journal of Experimental Psychology: General*, 144 (1), 114.

- Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2018), "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science*, 64 (3), 1155-1170.
- Dubinsky, Alan J. (1996), "Some Assumptions about the Effectiveness of Sales Training," *The Journal of Personal Selling and Sales Management*, 16 (3), 67-76.
- Eastwood, Joseph, Brent Snook, and Kirk Luther (2012), "What People Want From Their Professionals: Attitudes Toward Decision-Making Strategies," *Journal of Behavioral Decision Making*, 25 (5), 458-468.
- Edmondson, Amy C., Roderick M. Kramer, and Karen S. Cook (2004), "Psychological Safety, Trust, and Learning in Organizations: A Group-Level Lens," *Trust and distrust in organizations: Dilemmas and approaches*, 12, 239-272.
- Fiske, Susan T., and Shelley E. Taylor (1991), *Social Cognition*. McGraw-Hill Book Company.
- Frey, Carl Benedikt and Michael A. Osborne (2017), "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological forecasting and social change*, 114, 254-280.
- Hayes, Andrew F. (2013), "Serial Mediation Macro," (accessed August 11, 2020), <http://afhayes.com/spss-sas-and-r-macros-and-code.html>.
- Jackson, Dale E. (1988), *Interpersonal Communication for Technically Trained Managers: A Guide to Skills and Techniques*. New York: Quorum Books.
- Jacoby, Jacob (1974), "Consumer Reaction to Information Displays: Packaging and Advertising," in *Advertising and the Public Interest*, ed. S. F. Divita, Chicago: AMA, 101-118.
- Jacoby, Jacob, Donald E. Speller, and Carol Kohn (1974), "Brand Choice Behavior as a Function of Information Load," *Journal of Marketing Research*, 11, 63-69.
- Kannan, P. V., and Josh Bernoff (2019), "The Future of Customer Service Is AI-Human Collaboration," *MIT Sloan Management Review*, (May 29).
- Leung, Eugina, Gabriele Paolacci, and Stefano Puntoni (2018), "Man Versus Machine: Resisting Automation in Identity-Based Consumer Behavior," *Journal of Marketing Research*, 55 (6), 818-831.
- Li, Jia, Minghui Liu, and Xuan Liu (2016), "Why Do Employees Resist Knowledge Management Systems? an Empirical Study From the Status Quo Bias and Inertia Perspectives," *Computers in Human Behavior*, 65, 189-200.
- Logg, Jennifer M., Julia A. Minson, and Don A. Moore (2019), "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment," *Organizational Behavior and Human Decision Processes*, 151, 90-103.
- Longoni, Chiara, Andrea Bonezzi, and Carey K. Morewedge (2019), "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research*, 46 (4), 629-650.
- Luo, Xueming, Siliang Tong, Zheng Fang, and Zhe Qu (2019), "Frontiers: Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases," *Marketing Science*, 38 (6), 937-947.
- Martin, Warren S., and Ben H. Collins (1991), "Sales Technology Applications: Interactive Video Technology in Sales Training: A Case Study," *Journal of Personal Selling & Sales Management*, 11 (3), 61-66.
- Matheny, Meghan (2019), "Chorus.ai Announces Only Native Zoom Integration Delivering Most-Secure Recording Compliance for Enterprise," (accessed August 9, 2020), <https://www.businesswire.com/news/home/20190731005179/en/Chorus.ai-Announces-Native-Zoom-Integration-Delivering-Most-Secure>.

- McKinsey Global Institute (2018), "AI, Automation, and the Future of Work: Ten Things to Solve For," (accessed August 9, 2020), <https://www.mckinsey.com/featured-insights/future-of-work/ai-automation-and-the-future-of-work-ten-things-to-solve-for#>.
- Mende, Martin, Maura L. Scott, Jenny van Doorn, Dhruv Grewal, and Ilana Shanks (2019), "Service Robots Rising: How Humanoid Robots Influence Service Experiences and Elicit Compensatory Consumer Responses," *Journal of Marketing Research*, 56 (4), 535-556.
- Newell, Allen and Herbert Alexander Simon (1972), *Human Problem Solving*. Prentice-Hall Englewood Cliffs, NJ.
- Önköl, Dilek, Paul Goodwin, Mary Thomson, Sinan Gönöl, and Andrew Pollock (2009), "The Relative Influence of Advice From Human Experts and Statistical Methods on Forecast Adjustments," *Journal of Behavioral Decision Making*, 22 (4), 390-409.
- Puntoni, Stefano, Rebecca Walker Reczek, Markus Giesler, and Simona Botti (2020), "Consumers and Artificial Intelligence: An Experiential Perspective," *Journal of Marketing*, forthcoming.
- Ratneshwar, Srinivasan, and Shelly Chaiken (1991), "Comprehension's Role in Persuasion: The Case of Its Moderating Effect on the Persuasive Impact of Source Cues," *Journal of Consumer Research*, 18 (1), 52-62.
- Román, Sergio, Salvador Ruiz, and José Luis Munuera (2002), "The Effects of Sales Training on Salesforce Activity," *European Journal of Marketing*, 36 (11/12), 1344-1366.
- Shellenberger, Sue (2019), "The Robots That Manage the Managers," *Wall Street Journal* (April 15).
- Scammon, Debra L. (1977), "Information Load and Consumers," *Journal of Consumer Research*, 4, 148-155.
- Schultz, Mike (2011), "7 Reasons Why Sales Training Fails," (accessed August 16, 2020), <https://www.rainsalestraining.com/blog/7-reasons-sales-training-fails>.
- Simon, Herbert A. (1955), "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics*, 69 (1), 99-118.
- Slater, Stanley F., and John C. Narver (1995), "Market Orientation and the Learning Organization," *Journal of marketing*, 59 (3), 63-74.
- Srivastava, Sanjay (2019), "AI Adoption Accelerating, Yet Benefits at Risk Unless Businesses Address Customer and Employee Doubts," (accessed August 9, 2020), <https://www.itproportal.com/features/ai-adoption-accelerating-yet-benefits-at-risk-unless-businesses-address-customer-and-employee-doubts/>.
- Staelin, Richard and John W. Payne (1976), "Studies of the Information-Seeking Behavior of Consumers," in *Cognition and Social Behavior*, eds. John S. Carroll and John W. Payne, Hillsdale, NJ: Lawrence Erlbaum, 185-202.
- Sujan, Harish, Barton A. Weitz, and Nirmalya Kumar (1994), "Learning Orientation, Working Smart, and Effective Selling," *Journal of marketing*, 58(3), 39-52.
- Sun, Chenshuo, Zijun Shi, Xiao Liu, and Anindya Ghose (2019), "The Effect of Voice AI on Consumer Purchase and Search Behavior," Working paper.
- Sundblad, Willem (2018), "Data Is the Foundation for Artificial Intelligence and Machine Learning," *Forbes* (October 18).
- Teich, David A. (2019), "AI Chatbots: Companies Love Them; Consumers, Not So Much...," *Forbes* (April 9).
- Tversky, Amos, and Daniel Kahneman (1974), "Judgment Under Uncertainty: Heuristics and Biases," *Science*, 185 (4157), 1124-1131.

- Weitz, Barton A., Harish Sujan, and Mita Sujan (1986), "Knowledge, Motivation, and Adaptive Behavior: A Framework for Improving Selling Effectiveness," *Journal of marketing*, 50 (4), 174-191.
- Wilson, H. James, and Paul R. Daugherty (2018), "Collaborative Intelligence: Humans and AI Are Joining Forces," *Harvard Business Review*, 96 (4), 114-123.

Figure 1. An Overview of the Three Field Experiments

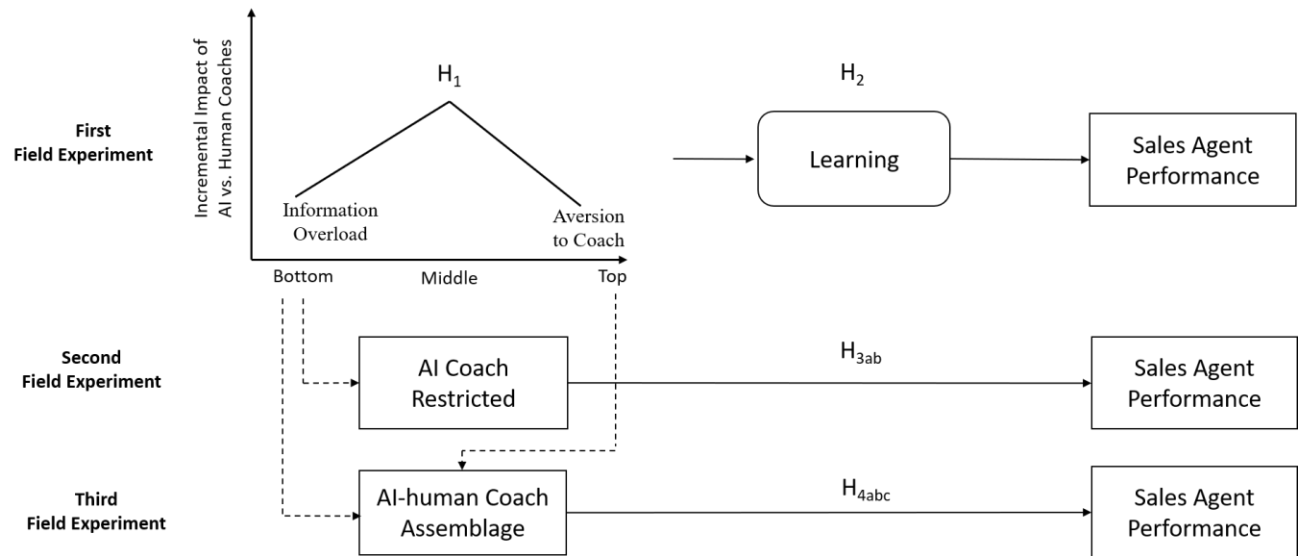
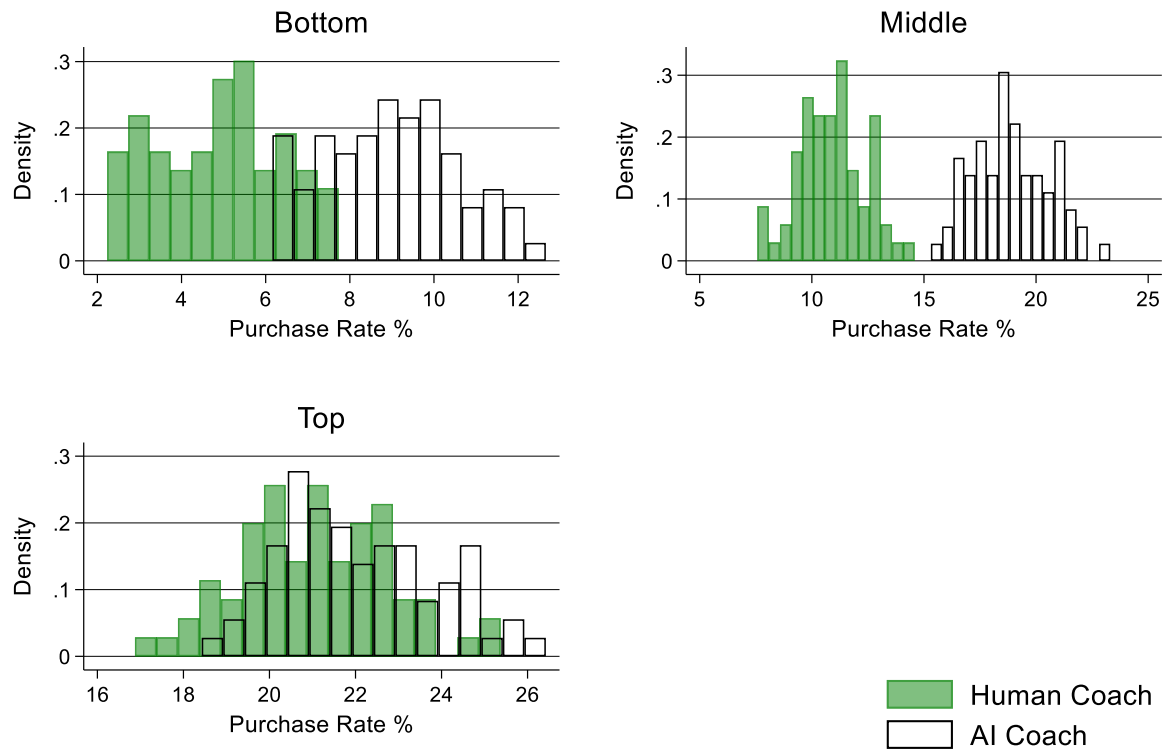
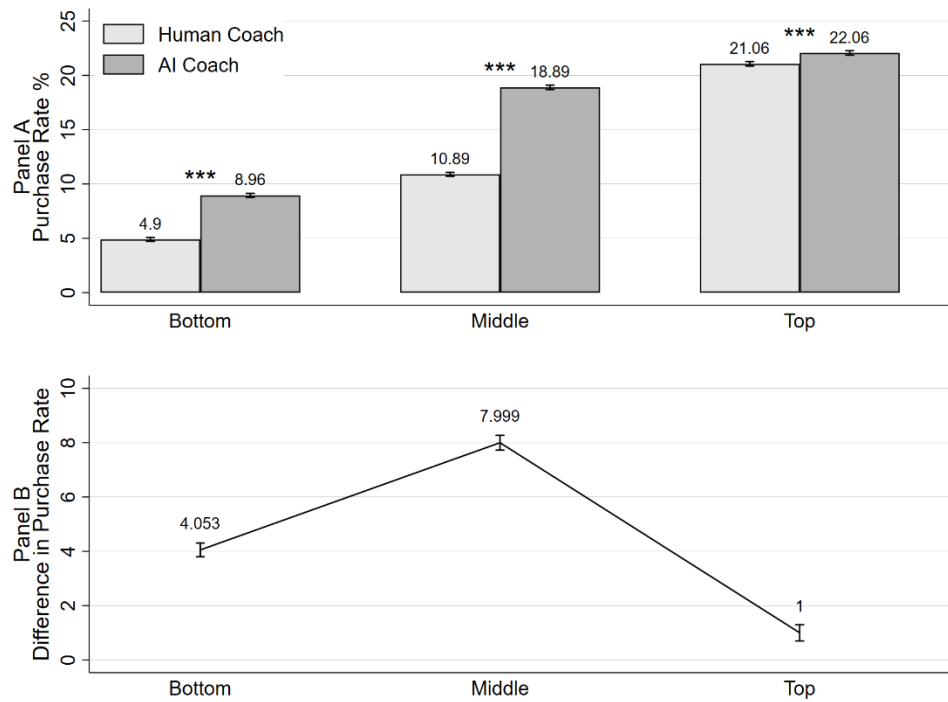


Figure 2. Model-Free Evidence of Sales Performance (Experiment 1)



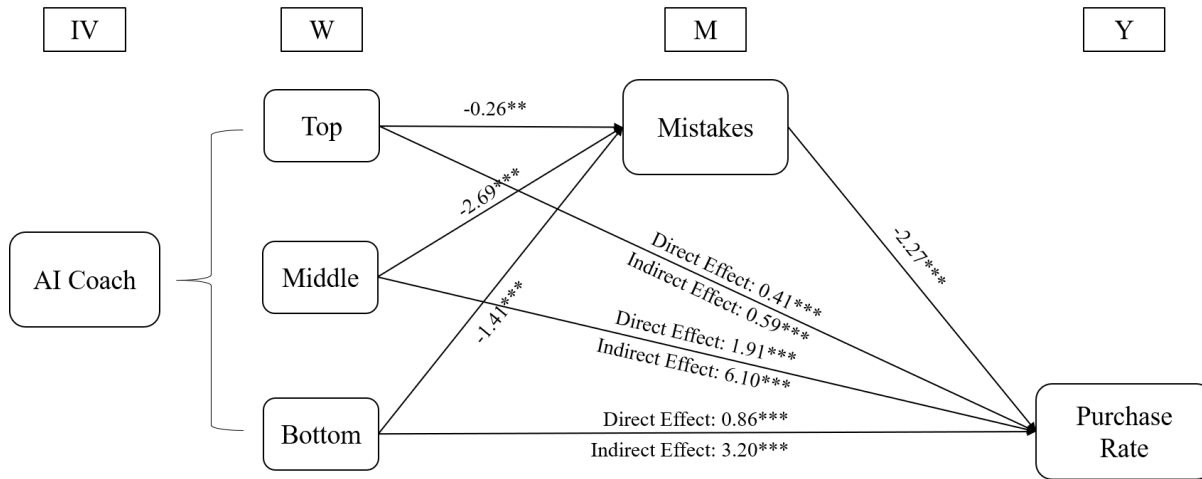
Notes: The width of the bins for all histograms is set to .5 percentage point.

Figure 3. Heterogeneous Effects of the AI versus Human Coach (Experiment 1)



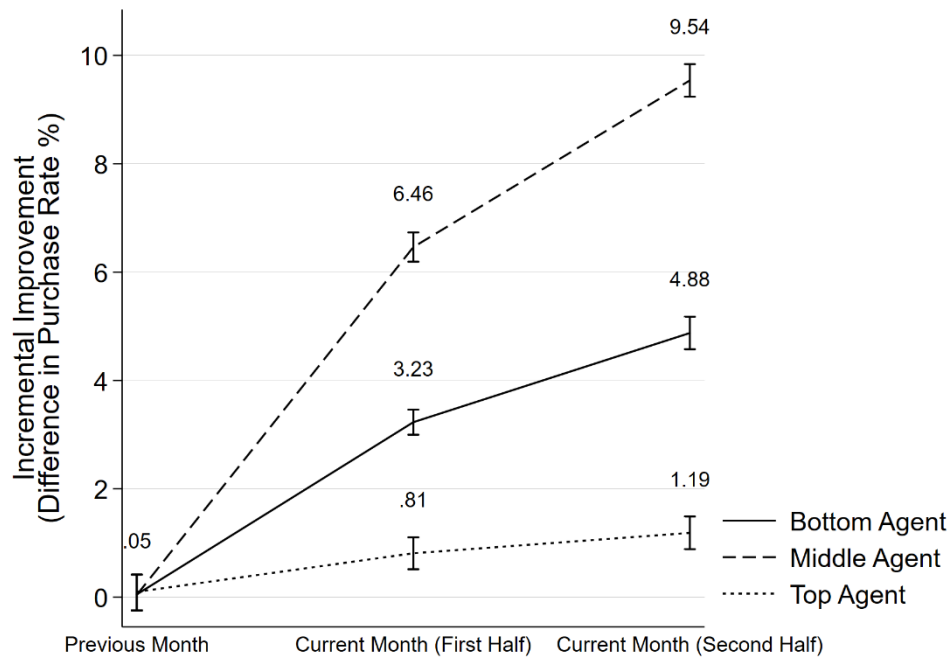
Notes: Error bars = +/- 1 SEs; *** = $p < .01$, ** = $p < .05$.

Figure 4. Moderated Mediation Results (Experiment 1)



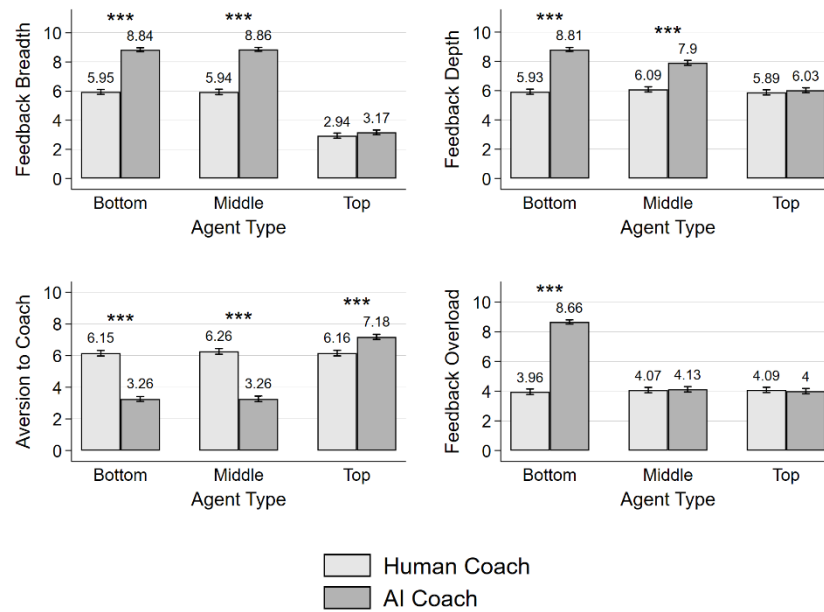
Notes: We use PROCESS Model 8 (Hayes 2013) with 5,000 bootstrap subsamples. *** = $p < .01$, ** = $p < .05$.

Figure 5. AI Training Learning Curves (Experiment 1)



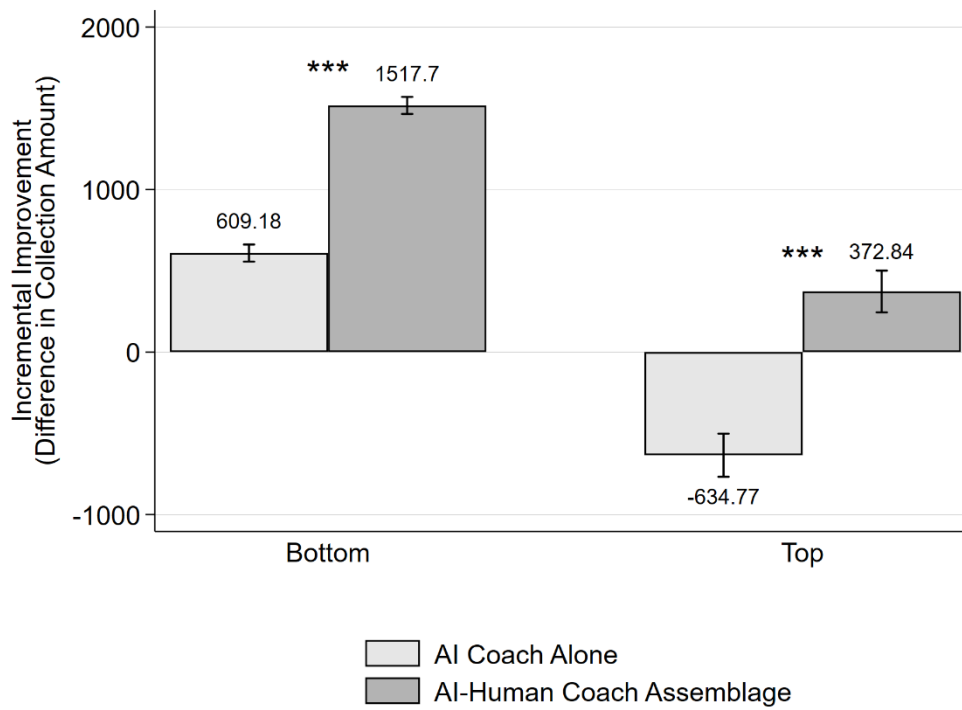
Notes: Each line represents the agents' incremental improvement with the AI vis-à-vis human coaches over different time periods. Error bars = +/- 1 SEs.

Figure 6. Sales Agents' Perceptions of the Coaching Feedback (Experiment 1)



Notes: Error bars = +/- 1 SEs; *** = p < .01, ** = p < .05.

Figure 7. Relative Effectiveness of the AI-human Coach Assemblage (Experiment 3)



Notes: The human coach group is the baseline. Notes: Error bars = +/- 1 SEs; *** = $p < .01$, ** = $p < .05$.

Table 1: Variable Definitions and Summary Statistics (Experiment 1)

Variable	Definition	Mean	S.D.	Min	Max
Purchase Rate (%)	The daily ratio of the sales calls converted into loan renewal to the total sales calls (averaged over the experiment month)	14.41	6.75	2.23	25.95
Purchase Rate_Half1 (%)	The purchase rate for the first half of the experiment (daily ratio averaged over the first 15 days in the experiment month)	13.97	6.67	2.27	25.75
Purchase Rate_Half2 (%)	The purchase rate for the second half of the experiment (daily ratio averaged over days 16 to 31 in the experiment month)	14.86	6.91	2.18	26.15
Purchase Rate_Previous (%)	The purchase rate for the month prior to the experiment	12.17	6.95	1.67	25.44
Bottom	Dummy variable for bottom-ranked agents whose performance was in the bottom 20% percentile in the month prior to the experiment	.34	.48	0	1
Middle	Dummy variable for middle-ranked agents whose performance was between 20th and 80th percentile in the month prior to the experiment	.33	.47	0	1
Top	Dummy variable for top-ranked agents whose performance was in the top 20% percentile in the month prior to the experiment	.33	.47	0	1
Calling Time	The daily active calling minutes averaged over the experiment month	143.81	27.16	105	188
Number of Calls	The daily attempted calls averaged over the experiment month	290.61	114.94	109	465
Mistakes (%)	The percentage of sales calls an agent made during the experiment that had at least one mistake identified	6.16	2.92	1	15.11
Agent Positive (%)	The percentage of sales calls an agent made during the experiment that had an overall positive sentiment of the agent	94.26	2.77	87	100
Customer Positive (%)	The percentage of sales calls an agent made during the experiment that had an overall positive sentiment of the customer	43.17	8.21	28	57
Customer Refuse (%)	The percentage of the customer's words of objection (such as no, don't, won't, etc.) in the sales call	20.5	4.59	11.9	30.1
Age	Age of the sales agent	21.99	2.27	18	26
Male	Dummy variable for male agents	.33	.47	0	1
High School	Dummy variable for agents who had high school as the highest degree	.41	.49	0	1
College	Dummy variable for agents who had college as the highest degree	.59	.49	0	1
Tenure <= 3 months	Dummy variable for agents whose tenure at the company was less than or equal to 3 months at the beginning of the experiment	.78	.41	0	1
Tenure 4-6 months	Dummy variable for agents whose tenure at the company was 4-6 months at the beginning of the experiment	.1	.31	0	1
Tenure > 6 months	Dummy variable for agents whose tenure at the company was more than 6 months at the beginning of the experiment	.11	.32	0	1
AI Experience	Dummy variable for sales agents who had the experience of using AI-powered personal assistants (e.g., Genie offered by Tmall.com, Xiao AI offered by the company Xiaomi), not the AI coach, at the beginning of the experiment	.05	.22	0	1
Feedback Breadth	The coach has comprehensively identified my sales mistakes in the training feedback (survey question with a 10-point Likert scale)	5.98	2.73	1	10
Feedback Depth	The coach has provided extensive suggestions for each mistake identified in the training feedback (survey question with a 10-point Likert scale)	6.79	1.82	4	10
Feedback Overload	The amount of information in the training feedback from the coach was too much for me to digest (survey question with a 10-point Likert scale)	5.36	2.11	1	9
Aversion to Coach	I was generally resistant to the training feedback from the coach (survey question with a 10-point Likert scale)	4.84	2.28	2	10

Table 2: Heterogeneous Impact of the AI Coach (Experiment 1)

	(1) Purchase Rate	(2) Purchase Rate (First Half)	(3) Purchase Rate (Second Half)
AI Coach	8.017*** (.280)	6.475*** (.272)	9.559*** (.309)
AI Coach * Bottom	-3.956*** (.389)	-3.243*** (.379)	-4.670*** (.429)
AI Coach * Top	-7.002*** (.394)	-5.649*** (.383)	-8.356*** (.435)
Bottom	-6.002*** (.282)	-5.962*** (.274)	-6.042*** (.311)
Top	10.162*** (.282)	10.183*** (.275)	10.140*** (.311)
Age	.029 (.036)	.020 (.035)	.038 (.040)
Male	-.102 (.171)	-.090 (.167)	-.114 (.189)
College	-.062 (.176)	-.023 (.172)	-.100 (.194)
Tenure 4-6 Months	-.215 (.271)	-.162 (.263)	-.268 (.298)
Tenure > 6 Months	.088 (.280)	.092 (.273)	.083 (.309)
AI Experience	.058 (.377)	-.075 (.366)	.190 (.415)
Constant	10.334*** (.816)	10.463*** (.794)	10.204*** (.899)
Observations	429	429	429
R2	.942	.944	.933

Notes: The dependent variable in Column 1 is the sales agent's purchase rate during the experiment month, while the dependent variables in Columns 2 and 3 are the sales agent's purchase rate for the first and second halves of the experiment month, respectively. Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 3: Additional Results (Experiment 1)

	(1) Calling Time	(2) # of Calls	(3) Agent Positive	(4) Customer Positive	(5) Customer Refuse	(6) Feedback Breadth	(7) Feedback Depth	(8) Aversion to Coach	(9) Feedback Overload
AI Coach	-4.035 (4.491)	10.860 (19.391)	3.616*** (.328)	10.226*** (.343)	-5.309*** (.289)	2.919*** (.232)	1.812*** (.238)	-2.959*** (.246)	.106 (.252)
AI Coach * Bottom	7.678 (6.243)	-17.265 (26.954)	-.172 (.456)	-4.368*** (.477)	2.723*** (.402)	-.012 (.322)	1.089*** (.331)	.077 (.342)	4.627*** (.350)
AI Coach * Top	-.812 (6.321)	33.860 (27.292)	-3.483*** (.462)	-9.015*** (.483)	4.693*** (.407)	-2.709*** (.326)	-1.691*** (.335)	3.984*** (.346)	-.234 (.354)
Bottom	12.882*** (4.524)	24.934 (19.531)	-2.024*** (.330)	-7.479*** (.346)	3.764*** (.291)	-.055 (.233)	-.163 (.240)	-.151 (.248)	-.123 (.253)
Top	1.280 (4.526)	-34.954* (19.539)	3.001*** (.330)	12.365*** (.346)	-6.648*** (.291)	-2.970*** (.233)	-.146 (.240)	-.096 (.248)	.029 (.254)
Age	-.529 (.577)	-.288 (2.490)	-.033 (.042)	.013 (.044)	-.015 (.037)	-.014 (.030)	.036 (.031)	.010 (.032)	.012 (.032)
Male	.137 (2.745)	.093 (11.852)	.185 (.200)	-.014 (.210)	-.116 (.177)	-.176 (.142)	-.087 (.146)	-.005 (.150)	-.071 (.154)
College	.815 (2.829)	-4.379 (12.213)	-.201 (.207)	-.146 (.216)	.252 (.182)	-.049 (.146)	-.184 (.150)	.212 (.155)	.355** (.158)
Tenure 4-6 Months	7.873* (4.339)	14.562 (18.735)	-.093 (.317)	-.288 (.332)	.081 (.279)	-.184 (.224)	.204 (.230)	-.154 (.238)	.059 (.243)
Tenure > 6 Months	6.843 (4.497)	-4.145 (19.417)	-.114 (.328)	-.011 (.344)	-.546* (.290)	-.185 (.232)	-.165 (.239)	-.218 (.246)	-.133 (.252)
AI Experience	2.140 (6.041)	-62.948** (26.080)	-.121 (.441)	-.060 (.462)	-.263 (.389)	.431 (.312)	.627* (.321)	-.059 (.331)	.272 (.338)
Constant	149.224*** (13.082)	296.469*** (56.479)	93.546*** (.955)	38.542*** (1.000)	23.147*** (.842)	6.369*** (.675)	5.390*** (.694)	5.955*** (.717)	3.619*** (.733)
Observations	429	429	429	429	429	429	429	429	429
R2	.081	.043	.528	.941	.867	.759	.421	.543	.592

Notes: The dependent variables in Columns 1-5 are, respectively, sales agents' average daily calling time, average daily number of attempted calls, proportion of calls with agent positive sentiment, proportion of calls with customer positive sentiment, and proportion of objection words identified in customers' conversations with the sales agents. The dependent variables in Columns 6-9 are, respectively, agents perceived feedback breadth, feedback depth, aversion to coach, and feedback overload. Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 4: Effect of AI Coach with Restricted Feedback on Bottom-ranked Agents (Experiment 2)

	(1) Purchase Rate	(2) Feedback Overload	(3) Purchase Rate
AI Coach (Restricted)	4.635*** (.362)	-6.505*** (.246)	2.810*** (1.037)
Feedback Overload			-.280* (.150)
Age	-.025 (.078)	.020 (.053)	-.019 (.077)
Male	.105 (.377)	.105 (.256)	.134 (.372)
College	.231 (.361)	-.142 (.246)	.191 (.357)
AI Experience	-1.434 (1.842)	.786 (1.252)	-1.213 (1.822)
Constant	9.523*** (1.742)	8.247*** (1.184)	11.837*** (2.117)
Observations	100	100	100
R2	.638	.883	.651

Notes: The dependent variable in Columns 1 and 3 is the sales agent's purchase rate during the experiment month, while the dependent variable in Column 2 is the agent's perception of feedback overload from the coach (survey data). Standard errors in parentheses, * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 5: The Relative Effectiveness of the AI-human Coach Assemblage in Loan Collection (Experiment 3)

	(1) Collection Amount (Whole Sample)	(2) Collection Amount (Bottom-ranked Agents)	(3) Collection Amount (Top-ranked Agents)	(4) Collection Amount (Whole Sample)
AI-human Coach Assemblage	925.724*** (76.747)	1504.899*** (53.766)	337.590*** (127.656)	1508.445*** (97.418)
AI Coach Alone	-47.198 (76.806)	595.724*** (54.342)	-684.267*** (126.220)	601.591*** (98.469)
AI Coach Alone * Top				-1274.465*** (137.773)
AI-human Coach Assemblage* Top				-1163.951*** (138.385)
Top	22403.151*** (182.072)			23178.734*** (180.375)
Age	40.163*** (15.341)	3.003 (16.385)	47.150** (21.767)	38.446*** (13.813)
Community College	-119.475 (111.816)	105.478 (239.693)	-129.738 (144.169)	-175.358* (100.794)
College	72.766 (68.224)	-2.371 (45.961)	192.679 (121.116)	94.697 (61.422)
Tenure = 3-6 Months	-31.990 (120.316)	-26.931 (62.382)		-8.935 (108.284)
Tenure = 7-9 Months	-214.509 (180.186)		-231.566 (164.488)	-229.821 (162.215)
Tenure > 12 Months	-239.109 (203.353)		-169.615 (198.780)	-179.185 (183.095)
AI Experience	-152.485 (127.346)	-123.576 (76.927)	-140.452 (271.110)	-100.711 (114.719)
Constant	11225.107*** (333.880)	11633.938*** (354.227)	33787.561*** (502.853)	10836.213*** (302.863)
Observations	451	224	227	451
R2	.997	.792	.261	.997

Note: The dependent variable in Columns 1 to 4 is the sales agent's average daily loan collection amount measured in local currency. Columns 2 and 3 perform the regression with only bottom- and top-ranked agents, respectively. Standard errors in parentheses * $p < .1$, ** $p < .05$, *** $p < .01$.