Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases

Xueming Luo
Temple University

Siliang Tong
Nanyang Business School, Nanyang Technological University

Zheng Fang
Business School, Sichuan University

Zhe Qu
Fudan University

Nanyang Business School Research Paper No. 20-33

Nanyang Business School
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
https://www.ntu.edu.sg/

Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases

Xueming Luo, Siliang Tong, Zheng Fang, and Zhe Qu

June 2019

Forthcoming *Marketing Science*

Acknowledgement. *The authors gratefully acknowledge the anonymous Fintech company for sponsoring the field experiment.*

Xueming Luo is the Charles Gilliland Distinguished Chair Professor of Marketing, MIS, and Strategy, Director and Founder of Global Institute for Artificial Intelligence& Business Analytics, Golbal Center for Big Data and Mobile Analytics, Fox School of Business, Temple University, luoxm@temple.edu. Siliang Tong is a PhD Student in Fox School of Business, Temple University, tug76173@temple.edu, Zheng Fang is Professor of Marketing at Sichuan University, 149281891@qq.com, and Zhe Qu is Associate Professor of Information Systems at Fudan University, quz@fudan.edu.cn.

Machines versus Humans: The Impact of AI Chatbot Disclosure on Customer Purchases

Abstract

Empowered by artificial intelligence (AI), chatbots are surging as new technologies with both business potentials and customer pushback. This study exploits field experiment data on over 6,200 customers who are randomized to receive highly structured outbound sales calls from chatbots or human workers. Results suggest that undisclosed chatbots are as effective as proficient workers and four times more effective than inexperienced workers in engendering customer purchases. However, a disclosure of chatbot identity before the machine-customer conversation reduces purchase rates by over 79.7%. Additional analyses find that these results are robust to non-response bias and hang-ups, and the chatbot disclosure substantially decreases call length. Exploration of the mechanisms reveals that when customers know the conversational partner is not a human, they are curt and purchase less because they perceive the disclosed bot as less knowledgeable and less empathetic. The negative disclosure effect seems to be driven by a subjective human perception against machines, despite the objective competence of AI chatbots. Fortunately, such negative impact can be mitigated by a late disclosure timing strategy and customer prior AI experience. These findings offer useful implications for chatbot applications, customer targeting, and advertising in conversational commerce.

Keywords: Artificial intelligence, chatbot, conversational commerce, new technology, disclosure

Introduction

Chatbots are a popular new technology with unprecedented business potentials, galvanized by AI and machine learning. Essentially, AI chatbots are computer programs that simulate human conversations through voice commands or text chats and serve as virtual assistants to users. Google Duplex, a ground-breaking application of AI chatbots, can make restaurant and haircut reservations over the phone, wherein people answering the call may not know they are engaging conversations with bots (Google AI 2018).

The market size of chatbots is expanding quickly, from $250 million in 2017 to over $1.34 billion in 2024 (ClickZ 2018). More than 21% of U.S adults and over 80% of Generation Z use voice/text bots for information search and shopping (MasterCard 2018). Many brands such as American Eagle Outfitters and Domino's Pizza have rolled out chatbots to take orders or recommend products, and major platforms such as Amazon, eBay, Facebook, and WeChat have adopted chatbots for conversational commerce (New York Times 2018).

AI chatbots can provide several unique business benefits. First, they automate customer services and facilitate firm-initiated communications. Chatbots are equipped with sophisticated speech recognition and natural language processing tools that enable them to understand complex and subtle dialogs and address consumer requests with depth, compassion, and even humor (Wilson et al. 2017). Moreover, chatbots can converse friendly with customers, because they don't have bad days and never get frustrated or tired like humans. In addition, they can easily scale up to handle a large volume of customer communications for call center businesses.

Despite such potential benefits for the supply side, a key challenge for AI chatbot applications is customer pushback from the demand side (Froehlich 2018). Customers may feel uncomfortable in talking with computer programs for personal needs or letting chatbots assist purchase decisions. That is , humans may prejudice that chatbots lack personal feeling and empathy, perceiving bots as less trustworthy with payment information and product recommendations (i.e., the uncanny valley feelings and algorithm aversion in Dietvorst et al. 2018; Forbes 2018).

Therefore, firms face a dilemma of disclosing the usage of AI chatbot technology to customers. On the one hand, if firms disclose the machine identity, they might not gain the full business value of AI chatbots due to customer pushback. On the other hand, customers have the right to know whether it is a bot or a human that handles their communications because of business ethics (Socialmediaweek 2018). Moreover, regulators are increasingly concerned about customer privacy protection and have encouraged companies to be transparent on chatbot applications during customer communications (FTC 2017).

Against this backdrop, we collaborate with a large financial service company to conduct a randomized field experiment on chatbot disclosure. The company randomly assigned 6,255 customers to receive highly structured outbound sales calls from chatbots or human workers. A novel part of our experiment design is to vary the disclosure of chatbots (no disclosure, disclosure before conversation, disclosure after conversation, or disclosure after decision), as well as human expertise (proficient or inexperienced workers).[1] This allows us to test the causal impact of chatbot disclosure on customer purchases and compare the performance of chatbots and human workers in the six-condition experiment.

Our data suggest that undisclosed chatbots are as effective as proficient workers and four times more effective than inexperienced workers in engendering customer purchases. However, the disclosure of chatbot machine identity before conversation reduces purchase rates by over

[1] Based on past six-month sales call performance in the company, proficient workers are among the top 20% (seasoned), while inexperienced workers or underdogs are among the bottom 20% (rookies).

79.7%. Our results are robust to various falsification checks and additional analyses with non-response bias and hang-ups. Also, compared with the condition of no disclosure, disclosure before conversation substantially reduces the call length.

Next, we test the behavioral mechanisms by augmenting the field experiment with survey data and voice-mining of the conversation records. The survey data support that when customers know the conversational partner is not a human, they are brusque and purchase less because they perceive the disclosed bot as less knowledgeable and less empathetic. However, voice-mining of the objective conversation records suggests that the undisclosed chatbot is competent in terms of knowledge and empathy. Thus, the negative chatbot disclosure effect seems to be driven by a subjective human perception against machines, despite the objective competence of AI chatbots.

Moreover, we explore various ways to mitigate the negative effect of chatbot disclosure on customer purchases. We find that such negative impact can be allayed by a late disclosure timing strategy and customer prior AI experience.

Our research makes several contributions. It provides the first field experiment evidence for the business value of emerging AI technology and challenges of chatbot applications. Our field data and voice-mining approaches not only reveal the negative impact of chatbot disclosure on customer purchases, but also shed light on the underlying mechanism. Our findings of the mitigated effects are non-trivial because they empower marketers to target certain customer segments for more optimal value of AI chatbot services. Also, brands can advertise the role of experiential learning so as to cultivate consumer trust in chatbots, i.e., from aversion to appreciation of bots.

More broadly speaking, we extend the discussion about machines versus humans. Our data suggest that undisclosed chatbots that incur almost zero marginal costs can outperform the paid underdogs by five times in purchase rates. These findings imply that the potential replacement of underperforming human workers by AI chatbots and other new automation technologies is an inevitable trend. However, our results of the negative disclosure effect also imply that chatbots may not perfectly substitute human labors in the near future due to a subjective human perception again bots. These findings have useful implications for chatbot applications in conversational commerce. Indeed, motivated by our findings, the financial service company has taken actions to implement a human-AI assemblage strategy. AI chatbots assist call center workers, especially the underdogs, by analyzing customer queries and emotional stress with voice-mining and by displaying best answers from the depository of company knowledge bank as possible solutions to customer needs.

Related Literature on AI Applications and Text-based bots

Prior research has recognized the benefits of AI technologies across various fields. In finance, trading bots and robo-advisors can facilitate investors for stock analytics (Trippi and Turban 1993). AI applications can improve banks' operation efficiency, fraud detection, and asset management (Fethi and Pasiouras 2010). Studies in healthcare have explored how AI powered algorithms can help doctors diagnose cancers (Esteva et al. 2017; Leachman and Merlino 2017). AI applications can reduce medical errors and improve hospital efficiency (Bennett and Hauser 2013; Patel et al. 2009). In marketing, Huang and Rust (2018) note that the future trend of AI applications hinges upon empathetic tasks that require computers to understand people's emotional status and respond appropriately with care and feeling. Leung, Paolacci, and Puntoni (2018) find that AI automation may be undesirable to consumers when the identity motives are important drivers of consumption. However, Logg et al. (2019) document that non-experts appreciate algorithmic advices based on lab experiments. Prior research also discusses how AI and robots will replace labor and work force (Brynjolfsson and Mitchell 2017; Lu, Rui, and Seidmann 2018). We extend this literature by providing real-world field experiment evidence for the negative impact of AI chatbot disclosure on customer purchases. We demonstrate the challenges of and harsh reactions to disclosed chatbots in outbound sales calls, although the bots can simulate human conversations in an intelligent and empathetic manner.

Our work on voice-based chatbots is related to and extends the literature on text-based chatbots (e.g. Köhler et al. 2011; Mimoun, Poncin, and Garnier 2017; Saad and Abida 2016; Sivaramakrishnan, Wan, and Tang 2007). Compared with text-based bots, voice-based bots offer more anthropomorphism in the humanized computer representations and richer interaction data such as voice pitch and tone beyond the narratives. Importantly, narratives only capture what is said, but miss how it is said (e.g., Do the conversation participants raise their voices suddenly? Or is there a frustration tone?). Extending prior literature on text-based chatbots, our research involves voice-mining analytics that can provide auditory cues of the sentiment and intent of the conversation participants. Also, extending prior research with surveys or lab studies measuring soft outcomes such as perceived fun and social presence, we conduct a field experiment addressing the hard metrics in terms of customer purchases. We further leverage deep learning methods of voice-mining and survey data to identify behavioral mechanisms that might account for the negative impact of disclosed bots on customer purchases.

Company Background and Experiment Settings

The randomized field experiment was conducted by a major internet-based financial service company in Asia (who wishes to be anonymous). In terms of types of business, this company offers various financial services such as personal loans, refinance, and equity investments to individual customers through its mobile app. Ranked among the top 20 in the Fintech Internet loan industry, this multi-billion-dollar company has over 23 million registered customers. Its customers are from all major provinces in the country (see Appendix C). In our experiment, the customers are borrowers who keep high credit score and have successfully repaid their loans to the company in the past 11 months for the 12 monthly installments. Since only one repayment is left, there is a sales opportunity of loan renewal. Most loans are in the amount between USD $800 to $2,500 for the purposes of purchasing electronic products such as smartphones, computers, and TVs. The company can assign chatbots or human workers in its call center to make the outbound sales calls. In order to boost responses, the service agents inform customers about a special promotional offer for renewing the loan. The promotion is a 24-hour limited time offer to waive the regular loan application processing fees if the customer decides to renew the loan with the same terms (loan amount, interests, and installments). All the outbound sales calls occur on a Tuesday afternoon from 2 p.m. to 4 p.m. during working hours of the day, and most customers would be at their workplace rather than home.

The company implements a sophisticated voice AI chatbot in its call center to provide timely customer services and improve the operational efficiency with lower labor costs. Unlike traditional rule-based systems that only handle simple inquiries with pre-recorded messages, the voice chatbot can conduct live and natural conversations with customers. The AI chatbot here is trained with the company's call center voice data to emulate the best-performing human workers in terms of understanding financial loan product features and deploying adaptive selling strategies (Churchill et al. 1985) in serving customers over the phone. The chatbot is applied to make highly structured outbound sales calls, because outbound calls have relatively standard conversation content for computers to handle. In the setting of structured outbound calls, without disclosure, customers would not realize the machine identity of the AI chatbot over the phone.[2] The chatbot in our experiment has an optimized female voice, i.e., with the most appealing pitch, tone, speed, and intonation to capture customer attention. The company uses a female voice because there is no significant difference between optimized female and male voices in call performance during pilot tests. Indeed, most chatbots (e.g., Alexa) in the industry adopt a female voice. Next, we present the field experiment design.

Field Experiment Design

In the field experiment, the company randomly assigned customers to receive a sales call from either human agents or AI chatbots. Each customer receives only one call and is randomized into one of the six experimental conditions in a between-subject design. Figure 1 (the top Panel) presents the six conditions and sample sizes.

The first condition is *underdogs* in the call center, i.e., unseasoned human workers whose past six-month call-reports performance is among the bottom 20 percentile. The second condition is *proficient workers*, i.e., experienced human agents whose past performance is among the top 20 percentile. The third condition is AI chatbot *without disclosure*. In this group, the chatbot initiates the sales call without revealing its machine identity. For these three conditions, the agents stars the call with a greeting statement: “Dear Customer, I am the service agent of the Company XYZ” prior to communicating the promotional deal to the customers.

[2] The AI platform that invents this chatbot conducted and passed the Turing test during chatbot developments. On the basis of a pilot test of 283 customers, the corporate partner company confirmed that without disclosure, 97% of its customers did not recognize the bot machine identity over the phone, because our setting here is a structured straightforward task—outbound sales calls in less than 2 minutes (the mean call length is less than 70 seconds).

The fourth condition is AI chatbot with *disclosure before conversation*. Here, the chatbot reveals its machine identity at the beginning of the conversation with the customer. The disclosure of chatbot identity is a simple statement: "Dear Customer, I am the AI voice chatbot of the Company XYZ" prior to communicating the same promotional deal. The fifth condition is AI chatbot with *disclosure after conversation*. In this group, the chatbot does not reveal its machine identity (with the same statement as in the fourth condition) until after communicating the promotional deal to customers but right before they decide whether or not to purchase. The sixth and final condition is AI chatbot with *disclosure after decision*, wherein the chatbot reveals its machine identity (also with the same statement as in the fourth condition) right after customers decide whether or not to purchase. [3]

All service agents across the six conditions follow the same sales call procedure as shown in Online Supplement Appendix A. Service agents first greet customers and appreciate their good repayment history before offering the special promotion deal over the phone. If customers are not interested, the agent will try to remedy the sales call by elaborating that the deal is designed for high-value customers and expires in 24 hours and by encouraging customers to review the promotion details on the mobile app.[4] However, if customers are interested in the promotion, the agent will ask follow-up questions about their changes in job as well as credit card balance. Customers are then asked to confirm whether or not to renew the loan. If customers agree to renew the loan, they need to log on the mobile app to sign the documents (99% of the

[3] Customers can reverse their purchase decisions and cancel the order after they know the bot machine identity. However, we do not find such case in our data. In addition, all agents and customers speak Mandarin rather than local dialects in our data.

[4] Customers who are not interested in the promotion would say "no" and terminate the call, thus leaving little opportunity for agents to remedy. Nevertheless, all experimental conditions follow this same protocol.

people who agreed to do so indeed followed through ultimately according to the company records). Examples of the call transcripts of the six experimental conditions can be found in the Appendix B, and audio examples of the AI chatbot used in our experiment are available online. In the data, making a purchase means that customers agree to renew their loans during the promotion period with the financial service company.

Data and Randomization Check

Figure 1 shows that there are a total of 6,255 attempted customers who are called by service agents. Out of these, 255 are non-responses (customers who may be too busy or have changed their contact numbers), and each condition has 1,000 responses to achieve the promotion goal with an automated replacement technique. Our proprietary dataset includes rich information about the customers. Table 1 summarizes the descriptive statistics. According to our data, 77.4% of the customers are males with an average age of 30.86, and most of them have a high school or higher degree. The statistics also indicate that targeted customers tend to be young working professionals who frequently use credit cards and engage in online shopping. They have on average 1.26 credit cards, US$1,843 credit card spending, and US$107 online spending in the past 30 days, as well as 10 online personal loan inquiries in the past 30 days. Their personal loan amount with the company is around US$2,017. We conducted randomization checks with these background variables. The results in Table 2 suggest that there is no significant difference among these variables across the 6 experimental conditions according to F-test statistics. Thus, the data passed the randomization check.

Effects of Chatbot Disclosure on Customer Purchases

The model-free results based on the raw data across the six treatments in Table 3 suggest that the condition of disclosure before conversation tends to have lower purchase rates, higher hang-up rates, and shorter call length.

Next, we apply econometric models to test the effects. Because we have randomized field experiment data to identify causal effects, our modeling analyses of purchase rates are straightforward. We develop a logit model, where the unobserved purchase likelihood is a logit function of the randomized conditions as below:

$$Purchase\ Likelihood_i = \frac{Exp(U_i)}{Exp(U_i)+1}$$

$$U_i = \alpha + \alpha_1 * Underdogs_i + \alpha_2 * Without\ Disclosure_i + \alpha_3 * Before\ Conversation_i + \alpha_4 * After\ Conversation_i + \alpha_5 * After\ Decision_i + \Gamma Controls_i + \varepsilon_i \quad (1)$$

where U_i denotes the latent utility of making a purchase, and the dependent variable of purchase is whether or not the customer has decided to renew the loan. The key independent variables are the six groups in our experiment, i.e., the five dummy variables with the proficient human agent group as the comparison baseline. $Controls_i$ is a vector of control variables with individual customer profiles, including Gender, Age, Education, Location dummies (see Appendix C for a frequency distribution of the 33 provinces), Number of Credit Cards, Online Loan Inquiries, Loan Amount, Credit Card Spending, Online Spending, as well as Customer Voice Pitch (which are derived from speech-to-text, Word2Vec, and Hierarchical Softmax Python tools, see Appendix D for details). Note that in the natural holdout case, without any sales call, the organic purchase rate is zero during the promotion period, because customers would not know the loan renewal opportunity without the sales calls. Thus, all effects on purchases here are incremental.

Table 4 Columns (1) to (3) report the results for all attempted calls. Across three models (Logit, Probit, and OLS), the results consistently suggest that relative to proficient human workers, disclosing the chatbot machine identity before the conversation statistically significantly reduces customer purchase rates ($p<.01$).

Besides the statistical significance, we present the magnitude of the effects in Figure 2. Compared to without disclosure, disclosure before conversation decreases customer purchase rates dramatically by 79.7% (from 0.237 to 0.048).

Robustness Checks with Falsification Tests

Our results are robust to various falsification checks. First, because the AI chatbot is trained by the calling records of the company's proficient workers, their performance should be similar. Results in Table 4 indeed support that the purchase rate of *no disclosure* is not significantly different from that of *proficient workers* ($p>.10$). This also rules out an alternative explanation that it might be the bad service quality of the chatbot itself rather than the act of disclosure that drives the negative effects. Also, the *underdogs* generate a significantly low purchase rate of .05 ($p<.01$). This makes sense because they are inexperienced rookies and unseasoned call center employees in the company. Still, they get some purchase results because of the exerted sales efforts. Moreover, we expect that the condition of *after decision* will not differ from the condition of *proficient workers*, because it is after the fact (customers have already made the decision of purchasing or not). This is confirmed by the insignificant coefficient of *after decision* in Table 4, thus passing another sanity or falsification check.

More Robustness Checks with Non-Responses and Hang-ups

First, we conducted additional analyses with possible non-response bias. Customers are randomized to receive the call, but not answer it. Thus, one possible concern is that customers may self-select to ignore the call and not purchase. That is, not all attempted calls are answered by customers because some customers cannot answer the phone (as this study was done during work hours), and others might have changed their contact numbers. As presented in Figure 1 the middle Panel, our data have a total of 255 non-responses with a response rate of 96% from attempted customers. This high response rate is not surprising because the targeted loan borrowers may fear missing out important loan update information from the lending company. More importantly, our data suggest that the non-response rates are almost evenly distributed among the six experiment groups, ranging from 3.5% to 5% as shown in Figure 1 and Table 2 last column. We also run the models after excluding the non-responses. Results in Table 4 Columns (4) to (6) confirm that all our main results are robust. Thus, possible selection effects due to non-responses cannot explain our results.

Moreover, we check our data regarding hang-ups (defined as the cases where customers terminate the call within five seconds right after knowing the bot machine identity). If customers terminate the call or hang up too early, they might not have indicated their purchase decisions. As reported in Figure 1 bottom Panel, there are a total of 608 hang-ups. The condition of *disclosure before conversation* had 563 cases (hang-ups without much interaction with the AI chatbot), and the condition of *disclosure after conversation* had 45 cases (hang-ups after the initial interaction with the AI chatbot). The rest four groups had zero hang-up case. We rerun the models after further excluding the hang-ups so as to scale the purchase rate {= number of 'yes' purchase decisions / (number of 'yes' purchase decisions + number of 'no' purchase decisions)}. Again, Table 4 Columns (7) to (9) confirm that all our main results are robust after accounting for hang-ups. We also check the robustness by measuring hang-ups within 4 seconds, 3 seconds, 2 seconds, and 1 second after the bot machine identity disclosure, and again, all results are robust across these different measures of hang-ups. These analyses of hang-ups due to chatbot disclosure motivated us to dive deeper by examining call length.

Additional Analyses with Call Length

One plausible explanation for our results is that when customers know the conversational partner is not a human, they tend to be curt (i.e., hang up abruptly or terminate early) and purchase less. If so, the call length in the disclosed chatbot condition should be significantly shorter than that of the undisclosed chatbot condition. This is confirmed by the Appendix D histograms of call length. Among the six experimental conditions, the case of chatbot identity disclosure before conversation has the shortest call length. We also run the models with call length as the dependent variable. Results in Table 5 with both OLS and Tobit models consistently support the negative and significant effect of *before conversation* on call length for the samples of attempted calls, excluding non-responses, and excluding hang-ups. However, these results cannot reveal the underlying psychological mechanisms, which are explored next.

Behavioral Mechanisms for the Negative Effects of Chatbot Disclosure

To understand the behavioral mechanism, we augment the field experiment with subjective data from post-call surveys, as well as objective voice data from audio analytics of the conversation records. The surveys poll all customers who have completed or hanged up the calls and ask their satisfaction with the service agent's knowledge level and sentimental empathy (See Appendix E). Figure 3 reports the results of a formal mediation test with 5,000 replications in bootstrapping (Preacher and Hayes 2004). The results confirm that relative to no disclosure, chatbot disclosure before conversation significantly reduces the *perceived* knowledge and empathy of chatbots and, through these two mediational routes, decreases call length and purchase rates (all path $p<.01$, see Appendix E for more details). In other words, when customers know the conversational partner is not a human, they are brusque and purchase less because they perceive the disclosed bot as less knowledgeable and less empathetic. However, voice-mining of the objective conversation records suggests that the undisclosed chatbot is indeed as competent as proficient workers in terms of knowledge and empathy (see Appendix F). Thus, the negative impact of chatbot disclosure may be driven by a subjective human perception against machines, despite the objective competence of AI chatbots.

Additional Checks on Deception Feeling and Order Cancellation

Another alternative explanation is customer feeling of deception. However, in the condition of disclosure before conversation, the customers are informed upfront about the chatbot machine identity, i.e., the disclosure is done immediately. Thus, it is more likely that customers' subjective perception against the chatbot rather than their feeling of deception drives the negative disclosure effect. Also, voice-mining of the conversation records failed to find words with strongly negative feelings across all experimental conditions, another evidence of no serious deception feeling. Moreover, according to the company records, there are no order cancellation or overt consumer complaints against the company in the conditions of chatbot identity disclosure after the experiment.

Strategies to Mitigate the Negative Effects of Chatbot Disclosure

Mitigation Strategy One. Results in Table 4 on the coefficient comparisons indicate that customer purchase rates significantly improve when the disclosure is delayed from before conversation, to after conversation, and to after decision (all $p<0.01$). Thus, more interactions with and experiential learning of chatbots may help allay the negative chatbot disclosure effect. In other words, as long as the chatbot identity is disclosed, regardless of before or after conversation, customer purchase rates are negatively affected. However, disclosing the bot identity after the conversation helps mitigate such negative impact. This is reasonable because the customer might form a good impression in the first one-minute interaction with the AI chatbot, which can help reduce their distrust of the chatbot.

Mitigation Strategy Two. We also explore how customers' prior AI experience can affect the negative effects of chatbot disclosure. The dataset provided by the company includes a binary variable that indicates whether a customer downloaded and used other AI apps on the smartphone (1= has at least one AI app with smart digital agents similar to Google Allo, ELSA Speak, Cortana, FaceApp, Edison Assistant, and 0 = otherwise). As shown in Table 6, prior experience with AI induces more customer purchases. More importantly, the coefficient of the interaction term $Prior\ AI\ Experience_i * Before\ Conversation_i$ is positive and significant ($p<.01$), suggesting that prior AI experience is helpful in reducing the negative disclosure effect.

Conclusion and Future Research

This research examines AI chatbots, a timely and managerially relevant topic. On the basis of a six-condition field experiment, it finds that the disclosure of chatbot machine identity reduces purchase rates substantially. Further analyses reveal that customers tend to purchase less and even terminate the calls early because they perceive the disclosed chatbot as less knowledgeable and empathetic.

Our setting of structured outbound calls is limited since the chatbot only engages in a restricted two-way information exchange rather than a highly interactive two-way conversation. This restrictive nature is an important limitation here, which may help open up new research. For example, it would be fruitful for future research to investigate dynamic differences of the two-way conversation between chatbot-customer dyads versus worker-customer dyads. Another direction for future research is to test the generalizability of our results in other settings such as the more dynamic inbound calls. Moreover, we address the first-order disclosure effects (with or without disclosure). Future research may test the second-order effects with different framings in the introduction of disclosed bots. For instance, the AI chatbots may self introduce to customers with the framing of enhanced technological benefits (big data computing and fast quantitative learning of AI chatbots), reduced customer hassle costs (less waiting time to get answers from AI chatbots), or even surprising consumer welfare (offering the product at a lower price because bots help save labor costs). Indeed, bots may help make life less prickly in certain interactions that are inherently bleak (e.g., call customer service support to fix computers or replenish a product). Paradoxically, in these interactions humans are trained to behave like a bot. Also, customers have different innate preferences of talking to bots, as some can be cordial and don't feel judged, but others tend to be rude and brusque (New York Times 2018). Thus, depending on the degree of task complexity and customer preference heterogeneity, future endeavors may let customers self-select who (bots or humans) to serve them over the phone in order to boost purchases in conversational commerce. As millions tell Alexa, Siri, or Google Assistant to play music, reorder products, and make appointments, the impact of AI new frontiers on our daily life will be ubiquitous in the long run.

In conclusion, more scholarly works are strongly encouraged to address this pivotal area of AI chatbot applications for marketing promotions and customer services.

Reference

Bennett, Casey C. and Kris Hauser (2013), “Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach,” *Artificial Intelligence in Medicine*, 57 (1), 9–19.

Brynjolfsson, Erik and Tom Mitchell (2017), “What can machine learning do? Workforce implications,” *Science*, 358 (6370), 1530–34.

Churchill, Gilbert A., Neil M. Ford, Steven W. Hartley, and Orville C. Walker (1985), “The Determinants of Salesperson Performance: A Meta-Analysis,” *Journal of Marketing Research*, 22 (2), 103.

ClickZ (2018), “Chatbot market size is set to exceed USD 1.34 billion by 2024 - ClickZ,” *clickz.com*.

Dietvorst, Berkeley J. Joseph P. Simmons, Cade Massey (2018) Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science* 64(3):1155-1170.

Esteva, Andre, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun (2017), “Dermatologist - level classification of skin cancer with deep neural networks,” *Nature*, 546 (7639), 115–18.

Fethi, Meryem Duygun and Fotios Pasiouras (2010), “Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey,” *European Journal of Operational Research*, 204 (2), 189–98.

Forbes (2018), “Conversational Commerce Is Where Online Shopping Was 15 Years Ago -- Can It Also Become Ubiquitous?,” *forbes.com*, (accessed August 3, 2018),

Froehlich, Andrew (2018), “Pros and Cons of Chatbots in the IT Helpdesk - InformationWeek,” *Informationweek.com*, (accessed October 18, 2018), [available at https://www.informationweek.com/strategic-cio/it-strategy/pros-and-cons-of-chatbots-in-the-it-helpdesk/a/d-id/1332942].

FTC (2017), “Privacy & Data Security Update (2016) | Federal Trade Commission,” *ftc.gov*, (accessed September 7, 2018), [available at https://www.ftc.gov/reports/privacy-data-security-update-2016].

Google AI (2018), “Google AI Blog: Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone,” *googleblog.com*, (accessed November 2, 2018), [available at https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html].

Huang, Ming-Hui and Roland T. Rust (2018), “Artificial Intelligence in Service,” *Journal of Service Research*, 21 (2), 109467051775245.

Köhler, Clemens F, Andrew J Rohm, Ko de Ruyter, and Martin Wetzels (2011), “Return on Interactivity: The Impact of Online Agents on Newcomer Adjustment,” *Journal of Marketing*, 75 (2), 93–108.

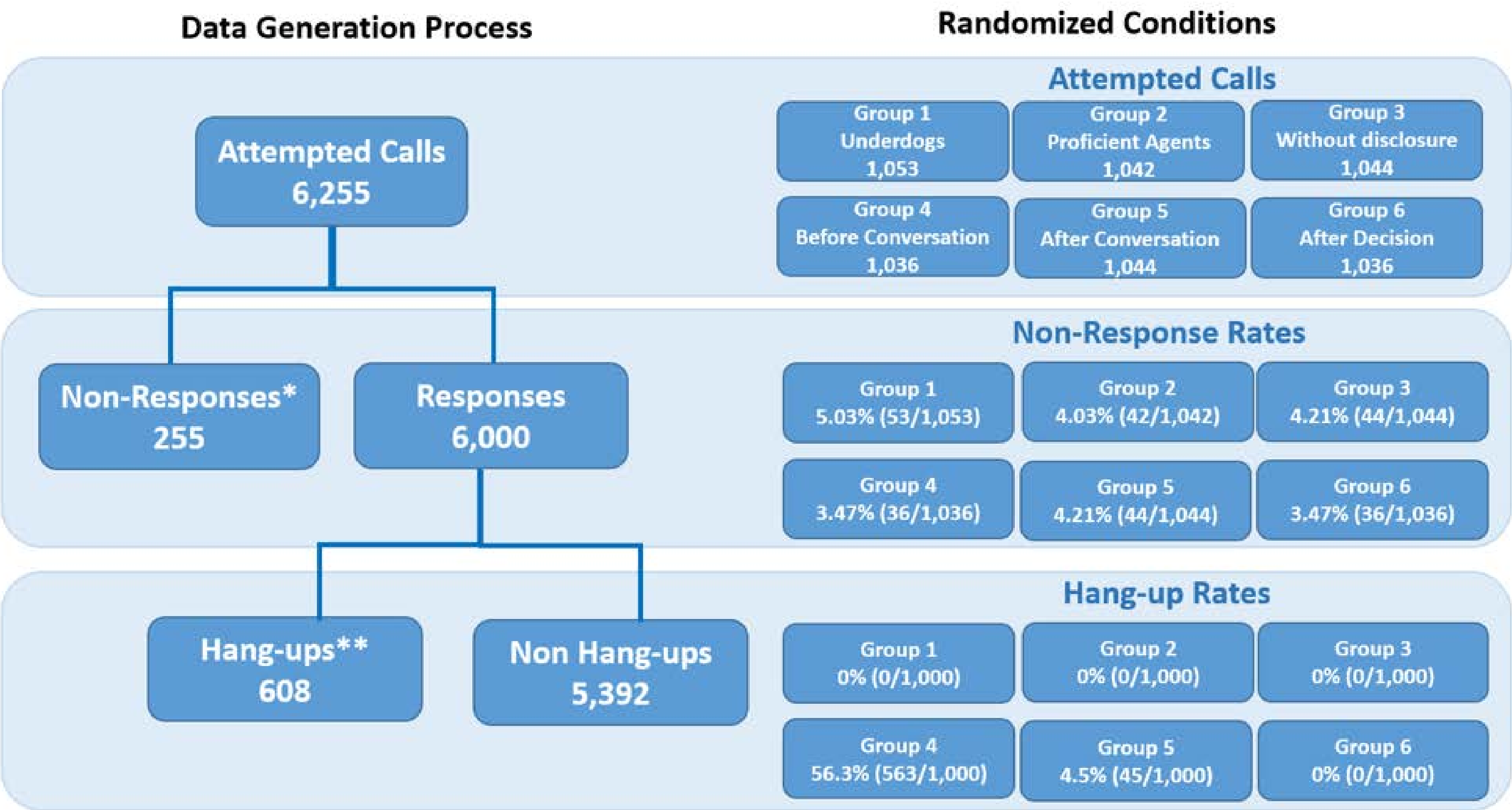
Leachman, Sancy A and Glenn Merlino (2017), “The final frontier in cancer diagnosis,” *Nature*, 542, 36.

Leung, Eugina, Gabriele Paolacci, and Stefano Puntoni (2018), “Man Versus Machine: Resisting Automation in Identity-Based Consumer Behavior,” *Journal of Marketing Research*, 55 (6), 818–31.

Logg, Jennifer M, Julia A Minson, and Don A Moore (2019), “Algorithm Appreciation : People Prefer Algorithmic To Human Judgment Algorithm Appreciation : People Prefer Algorithmic To Human Judgment,” *Organizational Behavior and Human Decision*

Processes, (151), 90–103.
Lu, Susan F., Huaxia Rui, and Abraham Seidmann (2018), “Does Technology Substitute for Nurses? Staffing Decisions in Nursing Homes,” *Management Science*, 64 (4), 1842–59.
MasterCard (2018), “Conversational commerce: A new opportunity for card payments.”
Mimoun, Ben, Mohammed Slim, Ingrid Poncin, and Marion Garnier (2017), “Animated conversational agents and e-consumer productivity: The roles of agents and individual characteristics,” *Information & Management*, 54 (5), 545–59.
New York Times (2018), “May A.I. Help You? - The New York Times,” *newyorktimes.com*, (accessed February 15, 2019), [available at https://www.nytimes.com/interactive/2018/11/14/magazine/tech-design-ai-chatbot.html].
Patel, Vimla L., Edward H. Shortliffe, Mario Stefanelli, Peter Szolovits, Michael R. Berthold, and Ameen Abu-Hanna (2009), “The coming of age of artificial intelligence in medicine,” *Artificial Intelligence in Medicine*, 46 (1), 5–17.
Piercy, Nigel F., David W. Cravens, and Neil A. Morgan (1999), “Relationships between Sales Management Control, Territory Design, Salesforce Performance and Sales Organization Effectiveness,” *British Journal of Management*, 10 (2), 95–111.
Saad, Ben, Sihem and Fatma Choura Abida (2016), “Social Interactivity and Its Impact on a User’s Approach Behavior in Commercial Web Sites: A Study Case of Virtual Agent Presence,” *Journal of Marketing Management*, 4 (2), 2333–6099.
Sivaramakrishnan, Subramanian, Fang Wan, and Zaiyong Tang (2007), “Giving an ‘e-human touch’ to e-tailing: The moderating roles of static information quantity and consumption motive in the effectiveness of an anthropomorphic information agent,” *Journal of Interactive Marketing*, 21 (1), 60–75.
Socialmediaweek (2018), “New Media Doesn’t Mean New Rules: The Challenges of Chatbots,” *socialmediweek.org*, (accessed September 7, 2018).
Trippi, Robert R. and Efraim. Turban (1993), *Neural networks in finance and investing : using artificial intelligence to improve real-world performance*, Probus Pub. Co.
Wager, Stefan and Susan Athey (2018), “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests,” *Journal of the American Statistical Association*, 113 (523), 1228–42.
Wilson, H James, Paul R Daugherty, and Nicola Morini-Bianzino (2017), “The Jobs That Artificial Intelligence Will Create,” *MIT Sloan Management Review*, 58 (4).

Figure 1: Experiment Design and Data Generation Process



Note: * Non-responses refer to the calls that were not answered by customers

** Hang-ups refers to the calls that were answered by customers, but the customers are terse and terminate the calls within five seconds right after chatbot identity disclosure

Figure 2: Purchase Rates across Experimental conditions

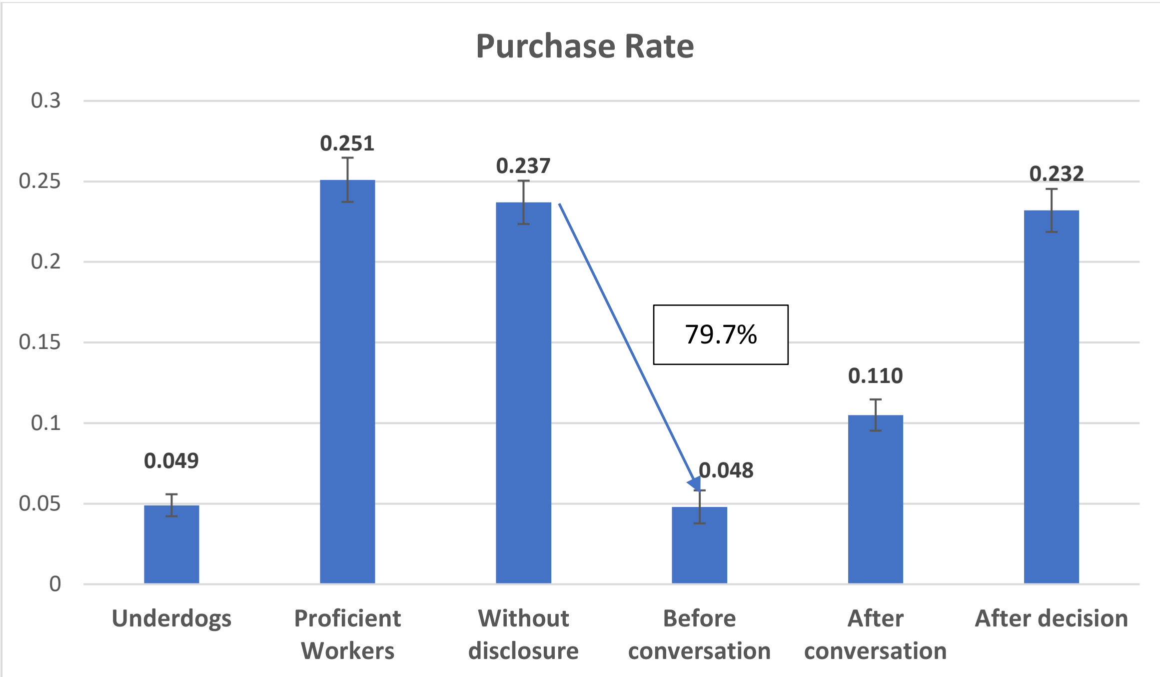
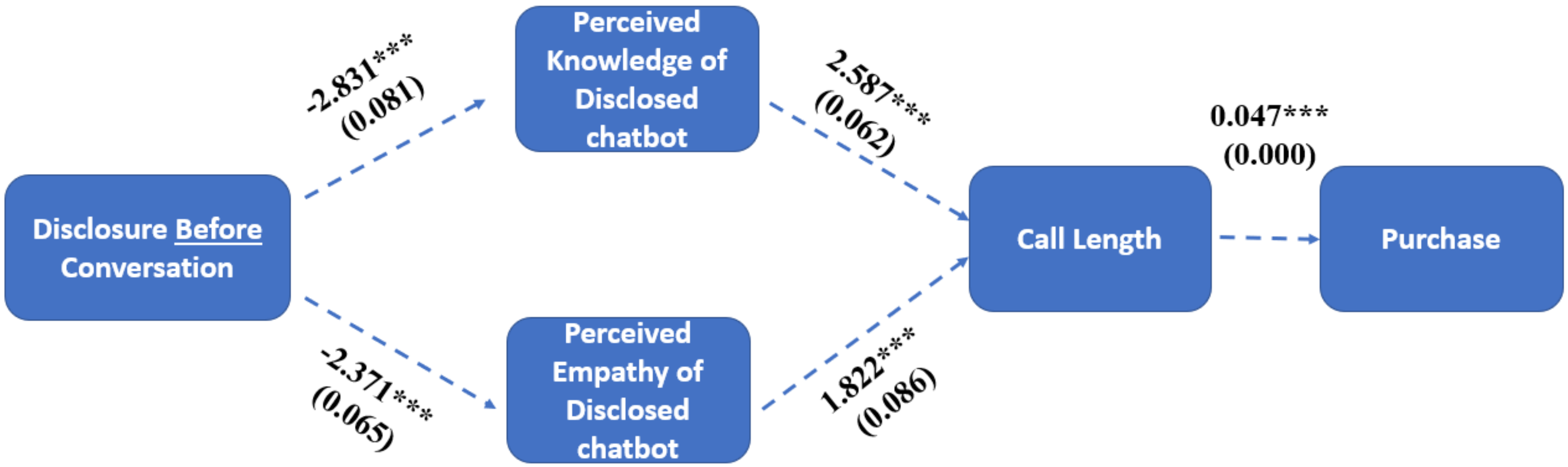


Figure 3: Customers are curt and purchase less from the disclosed chatbot, because they perceive the bot as less knowledgeable and empathetic



- Mediation test procedures recommended by Preacher and Hayes (2004) with 5,000 replications in bootstrapping

Table 1: Descriptive Statistics

Variables	Data Type	Explanations	Min	Max	Mean	25th Percentile	50th Percentile	75th Percentile	90th Percentile
Gender	Binary	1=Male, 0=Female	0	1	0.774	1	1	1	1
Age	Integer	The actual age calculated based on ID card information	19	55	30.86	26	30	34	40
Education	Category	1=Middle school and below, 2=High School, 3=Junior college or community college, 4=Undergraduate,	1	5	2.671	2	3	3	4
Number of Credit Cards	Count	Number of eligible credit cards owned by the customer	1	8	1.26	1	1	1	2
Online Loan Inquiries	Count	Number of personal loan inquies by the customer in the past 30 days	0	64	10	4	8	15	22
Loan Amount	Integer	The loan amount repaid by the customer to the company (US$)	142.857	2,857.143	2,017.049	1,285.714	2,142.857	2,857.143	2,857.143
Credit Card Spending	Continuous	Amount of credit card spending in the past 30 days (US$)	0	82,855	1,843	122.509	611.363	1,906.558	4,522.721
Online Spending	Continuous	Amount of online spending in the past 30 days (US$)	0	7415.617	107.337	11.938	36.206	92.347	227.716

Table 2: Randomization Check

Group	N	Gender	Age	Education	Number of Credit Cards	Online Loan Inquiries	Loan Amount	Credit Card Spending	Online Spending	Non-response Rate
Underdogs	1,053	0.759	30.854	2.696	1.247	10.877	2,035.502	1,867.649	115.176	5.03%
Proficient workers	1,042	0.788	30.750	2.663	1.287	10.242	2,040.951	1,863.102	97.539	4.03%
Without disclosure	1,044	0.778	30.921	2.677	1.295	10.139	1,984.907	1,993.738	115.803	4.21%
Before conversation	1,036	0.777	30.918	2.679	1.230	9.990	1,995.905	1,663.772	97.901	3.47%
After conversation	1,044	0.786	30.789	2.670	1.247	10.404	2,023.249	1,778.202	97.249	4.21%
After decision	1,036	0.769	30.911	2.667	1.236	10.433	2,036.715	2,070.709	117.982	3.47%
F-value		0.717	0.1332	0.2174	2.081	1.253	0.724	1.421	1.273	
P-value		0.610	0.985	0.955	0.065	0.281	0.605	0.213	0.272	

Table 3: Model-Free Results

Condition	N	Call Response Rate	Hang-up Rate	Call Length	Purchase Rate
Underdogs	1,053	94.96%	0.00%	39.888	0.049
Proficient Workers	1,042	95.97%	0.00%	63.888	0.251
Without disclosure	1,044	95.79%	0.00%	64.152	0.237
Before conversation	1,036	96.52%	56.30%	10.325	0.048
After conversation	1,044	95.78%	4.50%	63.873	0.110
After decision	1,036	96.52%	0.00%	63.731	0.232

Table 4: The Negative Disclosure Impact on Customer Purchases

DV: Purchase Rate		Attempted Calls			Excluding Non-Responses			Excluding Hang-ups	
	Logit	Probit	OLS	Logit	Probit	OLS	Logit	Probit	OLS
Underdogs (a)	-1.886***	-0.991***	-0.194***	-1.683***	-0.879***	-0.176***	-1.769***	-0.932***	-0.185***
	(0.164)	(0.080)	(0.015)	(0.217)	(0.113)	(0.023)	(0.219)	(0.114)	(0.025)
Without disclosure (b)	-0.082	-0.050	-0.014	-0.085	-0.051	-0.015	-0.085	-0.052	-0.015
	(0.104)	(0.061)	(0.015)	(0.105)	(0.062)	(0.016)	(0.105)	(0.062)	(0.016)
Before conversation (c)	-2.228***	-1.146***	-0.208***	-1.792***	-0.899***	-0.159***	-1.675***	-0.883***	-0.169***
	(0.187)	(0.087)	(0.015)	(0.366)	(0.195)	(0.041)	(0.379)	(0.201)	(0.044)
After conversation (d)	-1.056***	-0.585***	-0.141***	-1.065***	-0.595***	-0.146***	-1.017***	-0.570***	-0.142***
	(0.127)	(0.068)	(0.015)	(0.128)	(0.069)	(0.016)	(0.128)	(0.070)	(0.016)
After decision (e)	-0.107	-0.064	-0.018	-0.112	-0.068	-0.020	-0.113	-0.067	-0.020
	(0.105)	(0.061)	(0.015)	(0.106)	(0.062)	(0.016)	(0.106)	(0.062)	(0.016)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y
Excluding non-responses	N	N	N	Y	Y	Y	Y	Y	Y
Excluding hang-ups	N	N	N	N	N	N	Y	Y	Y
Constant	-0.741	-0.463	0.295***	-1.126	-1.601	-0.020	2.050	0.856	0.561
	(0.547)	(0.314)	(0.072)	(34.884)	(19.429)	(4.173)	(35.249)	(19.830)	(4.605)
N	6,255	6,255	6,255	6,000	6,000	6,000	5,392	5,392	5,392
Log likelihood	-2354.020	-2352.805	N/A	-2310.453	-2309.193	N/A	-2238.702	-2237.754	N/A
(Pseudo) R Squared	0.092	0.093	0.063	0.095	0.095	0.066	0.076	0.077	0.053
F-Value									
d-c	33.902***	36.546***	20.104***	3.787**	2.357	0.089	2.847*	2.320	0.363
e-d	54.784***	57.051***	66.179***	54.625***	56.996***	65.841***	48.965***	51.064***	54.744***

Note: *** P<.01, ** p<.05, * p<.10. This table tests the effect of chatbot identity disclosure on purchase rates for three different samples. Results from Columns (1) to (3) are based on the full sample of 6,255 attempted calls. Results from Columns (4) to (6) are based on the responded calls of 6,000 (excluding the 255 non-responses). Results from Columns (7) to (9) are based on the sample of 5,392 non hang-ups (further excluding the 608 hang-ups).

Table 5: The Negative Disclosure Impact on Call Length

DV: Call Length	Attempted Calls		Excluding Non-Responses		Excluding Hang-ups	
	OLS	Tobit	OLS	Tobit	OLS	Tobit
Underdogs (a)	-23.456***	-23.547***	-24.030***	-24.030***	-23.999***	-23.999***
	(0.546)	(0.567)	(0.283)	(0.282)	(0.283)	(0.282)
Without disclosure (b)	0.129	0.117	0.263	0.263	0.274	0.274
	(0.547)	(0.568)	(0.283)	(0.282)	(0.283)	(0.282)
Before conversation (c)	-51.373***	-51.543***	-53.566***	-53.566***	-49.447***	-49.447***
	(0.548)	(0.569)	(0.283)	(0.282)	(0.362)	(0.362)
After conversation (d)	-0.134	-0.143	-0.015	-0.015	0.603**	0.603**
	(0.548)	(0.568)	(0.283)	(0.282)	(0.286)	(0.285)
After decision (e)	0.238	0.255	-0.153	-0.153	-0.144	-0.144
	(0.548)	(0.568)	(0.283)	(0.282)	(0.283)	(0.282)
Control variables	Y	Y	Y	Y	Y	Y
Excluding non-	N	N	Y	Y	Y	Y
Excluding hang-ups	N	N	N	N	Y	Y
Constant	60.698***	60.406***	177.832**	177.832**	154.162*	154.162*
	(2.644)	(2.756)	(75.946)	(75.660)	(79.948)	(79.837)
N	6,255	6,255	6,000	6,000	5,392	5,392
Log likelihood	N/A	-24328.355	N/A	-19545.183	N/A	-17578.171
(Pseudo) R square	0.707	0.129	0.911	0.271	0.855	0.229
F-Value						
d-c	8743.801***	8153.974***	35765.645***	36035.915***	18834.990***	18887.533***
e-c	8831.520***	8243.613***	35591.725***	35860.680***	18528.711***	18580.399***

Note: *** P<.01, ** p<.05, * p<.10. This table tests the effect of chatbot identity disclosure on call length for three different samples. Results from Columns (1) to (3) are based on the full sample of 6,255 attempted calls. Results from Columns (4) to (6) are based on the responded calls of 6,000 (excluding the 255 non-responses). Results from Columns (7) to (9) are based on the sample of 5,392 non-hang-ups (further excluding the 608 hang-ups).

Table 6: The Negative Disclosure Impact on Customer Purchases is Mitigated by Prior AI Experience

Purchase Rate	Attempted Calls	Excluding Non-Responses	Excluding Hang-ups
Underdogs	-1.718***	-1.552***	-1.635***
	(0.196)	(0.245)	(0.247)
Without disclosure	-0.053	-0.058	-0.058
	(0.129)	(0.130)	(0.130)
Before conversation	-2.539***	-2.165***	-2.340***
	(0.270)	(0.428)	(0.505)
After conversation	-0.912***	-0.915***	-0.871***
	(0.154)	(0.154)	(0.154)
After decision	-0.003	-0.004	-0.004
	(0.128)	(0.128)	(0.128)
Prior AI experience	1.816***	1.817***	1.817***
	(0.177)	(0.177)	(0.177)
Underdogs*Prior AI experience	-0.247	-0.237	-0.244
	(0.386)	(0.386)	(0.386)
Without disclosure*Prior AI experience	0.183	0.188	0.183
	(0.262)	(0.262)	(0.262)
Before conversation*Prior AI experience	0.916**	0.900**	1.257**
	(0.405)	(0.405)	(0.533)
After conversation*Prior AI experience	-0.311	-0.310	-0.264
	(0.301)	(0.302)	(0.304)
After decision*Prior AI experience	-0.154	-0.146	-0.147
	(0.260)	(0.260)	(0.260)
Control variables	Y	Y	Y
Excluding non-responses	N	Y	Y
Excluding hang-ups	N	N	Y
Constant	-0.991*	-34.405	-29.953
	(0.580)	(36.974)	(37.361)
N	6,255	6,000	5,392
Log likelihood	-2116.459	-2115.005	-2048.078
Pseudo R-squared	0.171	0.171	0.155

Note: *** P<.01, ** p<.05, * p<.10.