Received: 11 February 2021 | Accepted: 27 June 2022

DOI: 10.1111/poms.13953

POMS

ORIGINAL ARTICLE

Voice-based AI in call center customer service: A natural field experiment

Lingli Wang[1] | Ni Huang[2] | Yili Hong[3] | Luning Liu[4] | Xunhua Guo[5] | Guoqing Chen[6]

[1]School of Modern Post, Beijing University of Posts and Telecommunications, Beijing, China

[2]Miami Herbert Business School, University of Miami, Coral Gables, Florida, USA

[3]Miami Herbert Business School, University of Miami, Coral Gables, Florida, USA

[4]School of Economics and Management, Harbin Institute of Technology, Harbin, China

[5]School of Economics and Management, Tsinghua University, Beijing, China

[6]School of Economics and Management, Tsinghua University, Beijing, China

Correspondence
Luning Liu, School of Economics and Management, Harbin Institute of Technology, Nangang District, Harbin 150001, China.
Email: liuluning@hit.edu.cn

Handling Editor: Subodha Kumar

Funding information
Beijing University of Posts and Telecommunications Basic Scientific Research Program, Grant/Award Number: 2022RC21; National Natural Science Foundation of China, Grant/Award Number: 72201038; Luning Liu acknowledges support from the National Natural Science Foundation of China, Grant/Award Numbers: 72034001, 71974044; Tsinghua University Initiative Scientific Research Program, Grant/Award Number: 2019THZWYX08; MOE Project of Key Research Institute of Humanities and Social Sciences at Universities, Grant/Award Number: 17JJD630006

Abstract

Voice-based artificial intelligence (AI) systems have been recently deployed to replace traditional interactive voice response (IVR) systems in call center customer service. However, there is little evidence that sheds light on how the implementation of AI systems impacts customer behavior, as well as AI systems' effects on call center customer service performance. By leveraging the proprietary data obtained from a natural field experiment in a large telecommunication company, we examine how the introduction of a voice-based AI system affects call length, customers' demand for human service, and customer complaints in call center customer service. We find that the implementation of the AI system temporarily increases the duration of machine service and customers' demand for human service; however, it persistently reduces customer complaints. Furthermore, our results reveal interesting heterogeneity in the effectiveness of the voice-based AI system. For relatively simple service requests, the AI system reduces customer complaints for both experienced and inexperienced customers. However, for complex requests, customers appear to learn from the prior experience of interacting with the AI system, which leads to fewer complaints. Moreover, the AI-based system has a significantly larger effect on reducing customer complaints for older and female customers as well as for customers who have had extensive experience using the IVR system. Finally, we find that speech-recognition failures in customer-AI interactions lead to increases in customers' demand for human service and customer complaints. The results from this study provide implications for the implementation of an AI system in call center operations.

KEYWORDS

artificial intelligence, customer service, difference-in-differences, natural field experiment, service flexibility

1 | INTRODUCTION

Advances in machine learning (ML) technology have accelerated the application of voice-based artificial intelligence (AI) systems in various business functions, performing tasks such as speech recognition and natural language processing.[1] With the intention of improving customer experience as well as reducing service costs, an increasing number of companies are deploying voice-based AI to complement or replace current systems and services provided by human agents (Xiao & Kumar, 2021). According to Markets and Markets (2021), the global conversational AI market size is predicted to grow from $6.8 billion in 2021 to $18.4 billion by 2026, and AI-supported customer service is a major factor driving the growth. Moreover, the value of global call center AI market reached $959.80 million in 2020 and is predicted to reach $9,949.61 million by 2030 (Valuates Reports, 2022).

Accepted by Subodha Kumar, after two revisions.

TABLE 1 Voice-based artificial intelligence (AI) system versus interactive voice response (IVR) system

	Voice-based AI system	IVR system
Inputs for Building the System	Large amounts of service data	Expert knowledge
Technological Characteristics	Natural language processing Speech recognition Improves with service data	Pre-designed services transferring rules Remains the same as originally designed
Customer-System Interaction	Interacting in natural dialogues	Inputting specific information
Service Organization	Direct routing	Hierarchical structure

Our study examines the implementation of a voice-based AI system that replaces the traditional interactive voice response (IVR) system in a customer service call center. In the absence of the AI system, customer calls are first connected to the IVR system, and customers communicate with the IVR system through phone keypads to obtain specific services. The service requests that the IVR system cannot handle are then transferred to human agents. Upon the rollout of a voice-based AI system, customers communicate with the AI in natural dialogues, and the AI system performs tasks, such as processing natural language in a manner that resembles human intelligence. Note that the voice-based AI system is different from the traditional IVR system in several ways as summarized in Table 1. To begin with, the AI system continuously evolves with the accumulation of a large amount of service data, enhancement in computing power, and improvement in learning algorithms (LeCun et al., 2015). In contrast, the IVR system was designed by industry experts based on their service experiences, does not change with service data, and requires customers to strictly follow pre-set rules when interacting with the system (Resnick & Virzi, 1995). Through speech recognition in the AI system, customers can tell the system which kinds of services they require; on this basis, they are directly routed to certain services (Tang et al., 2003). In contrast, the IVR system typically relies on a hierarchical structure that directs customers in a step-by-step manner to locate specific services (Suhm et al., 2002). To switch services, customers are required to return to the main service menu and repeat the above-mentioned actions to select another service.

Considering the technological advantages of the voice-based AI system, the implementation of such a system as a replacement for the traditional IVR system might significantly influence customer service experiences. First, the AI system improves the flexibility of service flows and enables personalized customer service. Instead of strictly following pre-defined service flows like in the IVR system, customers can actively control the pace of service when they interact with an AI-based system. For example, because of the AI system's flexible navigation structure, customers can skip over the layers of IVR structures and directly access the desired services. Second, the voice-based AI system can adapt to customers' interaction preferences to improve their service experiences. While the IVR system provides highly structured and limited choices for customers, the AI system can interact with customers in natural dialogues, allowing customers to express their needs adequately (Fountain et al., 2019). Moreover, the voice-based AI system has the ability to learn from prior interactions with the input data from customers and iteratively improve its performance. In scenarios wherein the AI system gets stuck, it can tag the problems with the help of human agents and learn from the scenarios to resolve similar problems in the future (Wilson & Daugherty, 2018).

According to Forbes Insights, call centers are predicted to be the new sandbox for AI-powered customer experience, as they are expected to deploy AI-based tools to boost retention, loyalty, and profit.[2] AI technology has gained widespread adoption in call center services over the past few years.[3] Nonetheless, despite the growing interests in AI, its implementation has a proof-of-concept-to-production gap (Perry, 2021). In other words, AI can work well theoretically or on test data, but it may fail to reach expectations in practical settings. In our study context, an IVR system operates with pre-designed fixed logic, while an AI-based system relies on a complex algorithmic structure. In turn, given the high variability of customer interactions in call center service, the service efficacy of an AI system is likely to be subject to variation (Brynjolfsson & Mcafee, 2017). For example, customers may speak with accents or dialects while communicating with AI, thereby resulting in speech-recognition failures that influence the effectiveness of the systems. However, without the rigid logic of an IVR system, an AI-based system might learn to adapt to the customer interactions and even be more effective than an IVR system. The high variability of tasks within customer service requires more flexible, human-like responses, and the AI system may work better than the strictly programmed IVR systems that are inflexible. Therefore, we believe it is important for researchers to empirically test the effectiveness of AI in real-life settings.

Most prior studies on AI in operation management (OM) have mainly focused on the effects of AI-supported automation and smartness and examine how related technologies are deployed to facilitate operation decisions or redesign operation process in product pricing (Karlinsky-Shichor & Netzer, 2019), order decision-making (Li & Li, 2022), and quality management (Senoner et al., 2021). Limited work explores the role of AI in interactions between customers and service systems, particularly in service contact design scenarios (A. V. Roth & Mentor, 2003). With few exceptions, for example, Cui et al. (2021) examined the role of AI in buyer price requests and its influence on seller price quotations in business-to-business wholesaling. Contributing to this knowledge gap in prior literature, our study focuses on the effects of implementing a voice-based AI system in the business-to-consumer customer service setting. Specifically, by analyzing data from a natural field experiment (NFE) in a large telecommunication company's customer service call center, we seek to answer the following research questions:

How does the introduction of voice-based AI systems impact call length, customer's demand for human service, and customer complaints in call center customer services? How do the effects of AI implementation in customer service vary for different customers?

To this end, we examine a NFE[4] with a voice-based AI system implemented in a telecommunication company's call center customer service operation. In the experiment, the company's customer service operation rolls out the AI system to replace the IVR system in different phases, serving a portion of its customers based on the last digit of the customer's phone number, thereby allowing them to engage in customer service calls through natural dialogues with AI. We then use difference-in-differences (DID) estimations to identify the effects of the AI-based system on key outcomes. Our results reveal that the duration of machine service and customers' demand for human service increases temporarily after the introduction of the voice-based AI system, suggesting a possible novelty effect. Meanwhile, the AI system significantly and persistently reduces customer complaints.

Moreover, we find interesting heterogeneity in the main effects of the AI system. To begin with, the effects of the AI system on customer complaints appear to depend on the complexity of the service requests. Compared with the customers who continue to use the traditional IVR system, the customers assigned to use the AI-based system tend to make fewer complaints when they have relatively simple service requests. In contrast, with relatively complex service requests (i.e., service calls transferred to human agents), customers learn from their prior interactions with AI; this learning effect leads to fewer complaints. Last, we find that the AI-based system exerts a significantly greater effect on reducing customer complaints for older customers, female customers, and for customers with longer user tenure.

Our study makes several important contributions to the related literature on AI applications and call center operations. First, our study adds to research on AI applications by extending the scope to the call center customer service setting and offers useful insights into how AI-powered service flexibility impacts different outcomes in human-AI interactions (Cui et al., 2021; Luo et al., 2019; Sun et al., 2019). Second, we contribute to the literature on call center customer service operations by empirically examining how the implementation of the AI system affects customer behavior and the performance of customer service, responding to calls for research on using disruptive technologies like AI to address OM problems in general (Karlinsky-Shichor & Netzer, 2019; Kumar et al., 2018) and to explore the direct effects of technology-mediated customer-involved service contact designs in particular (A. V. Roth & Mentor, 2003). In addition, building on related OM literature on call center operations, which views customers as mostly homogenous and uses a single metric to represent the performance of operation systems for all customers (Khudyakov et al., 2010; Tezcan & Behzad, 2012), our study further explores customer heterogeneity in responding to the operations of the voice-based AI system.

Furthermore, our findings also offer useful implications for practice. We demonstrate that using the voice-based AI system to replace the IVR system does not result in customer aversion to the AI system, thereby validating the effectiveness of using voice-based AI systems in call center customer service. We also demonstrate the novelty effect of implementing an AI system and find significant heterogeneity in the effectiveness of voice-based AI systems in reducing customer complaints based on the complexity of customer requests as well as customers' age, gender, and tenure with the company service. These results provide actionable insights into the implementation and further development of voice-based AI systems. For example, companies must consider the possible short-term increases in the duration of machine service and customers' demand for human service while scheduling a service system that applies voice-based AI to replace the IVR system. Instead of relying on customers' self-learning, companies could educate their customers on using an AI system with relatively complex requests. Moreover, the details obtained from customer-AI conversations reveal that speech-recognition failures may lead to negative consequences. Therefore, it is necessary for companies to continuously improve the capability of their AI systems to cater to a diverse customer base.

2 | RELATED LITERATURE

2.1 | Application of AI systems

Following prior literature, we define AI systems as algorithms that perform perceptual, cognitive, and conversational functions typical of the human mind (Longoni et al., 2019). In recent years, the significant development of AI systems has led to wide adoptions and applications in various domains. Specifically, in the OM literature (see the summary of related literature in Supporting Information A), from the technical perspective, some prior work attempted to design AI-based algorithms to solve operational problems such as demand or sales forecasting (Cui et al., 2018), product pricing (Yang et al., 2022), and quality inferring (Senoner et al., 2021). Meanwhile, scholars have also explored how AI-enabled automation and smartness features facilitate or support operational decisions in contexts such as price request (Cui et al., 2021), order decision-making (Li & Li, 2022), and automated pricing (Karlinsky-Shichor & Netzer, 2019).

Recently, a few studies on the application of AI systems have begun to understand the use of such systems in commerce operations, where the AI system directly interact with individuals. For example, Cui et al. (2021) examined how

AI chatbots affect suppliers' price quoting strategies. They found that automation of chatbots alone leads to discrimination against chatbot buyers, but signaling the use of a smart recommendation enabled by an AI system effectively reduces suppliers' price quote for chatbot buyers. In addition, Sun et al. (2019) suggested that the use of voice-based AI in online shopping significantly affects consumers' search behavior and purchase decisions. These prior studies primarily focused on the role of AI systems in facilitating sales, but not much is known regarding the application of voice-based AI systems in post-sales—that is, the scenario of customer service. In this regard, our study aims at addressing this research void in the stream of work on AI system applications by empirically examining the impact of implementing a voice-based AI system as a replacement for an IVR system for customer service on the key outcomes related to customer experience and call center operations.

2.2 | Information technology and service operation

Information technology plays an important role in improving service operations (A. V. Roth & Mentor, 2003). Companies increasingly rely on technology-based services to reduce service costs (Krishnan et al., 1999) and increase service efficiency (Beckman & Sinha, 2005). In recent years, with the accumulation of large amounts of data on customers and transactions, companies have gradually applied data-driven algorithms to automatically process service-related tasks—such as customer segmentation, pattern identification, service instruction, and real-time personalization—which, in turn, help companies to improve service quality (Sodhi et al., 2022; Sun et al., 2019).

Customers play an essential role in the delivery of services (A. V. Roth & Mentor, 2003). The design of customer contact—the interaction between a customer and a service provider—is important for shaping customers' service experiences (Kellogg & Chase, 1995). Prior research has empirically examined the contact between customers and employees (Kellogg & Chase, 1995; Soteriou & Chase, 1998) and demonstrated that the physical service environment significantly influences customers' perceptions and behavior (Bitner, 1992). However, the advancement of information technology is changing the ways in which customers interface with service providers. For example, companies commonly establish self-service systems to cater to customers' real-time service needs (Tezcan & Behzad, 2012). More recently, AI is being implemented to replace or complement conventional service providers (Xiao & Kumar, 2021). Froehle and Roth (2004) extended the customer contact perspective to technology-mediated services and called for research on exploring the effects of virtual service contact designs. Therefore, this paper focuses on the effects of different virtual service contact designs (i.e., IVR and AI systems) in the context of call center customer service and specifically investigates how replacing IVR systems with voice-based AI directly influences customers' interaction outcomes.

2.3 | Information technology and call center customer service

Call center customer service has been an essential channel through which customers interact with firms (Aksin et al., 2007; Tezcan & Behzad, 2012). New developments in information technology provide an opportunity to redesign and improve service-delivery operations in call centers. For example, information technology supports a call center to expand to a larger scale (Adria & Chowdhury, 2004). In such contexts, researchers have examined the effects of call center centralization (Adria & Chowdhury, 2004) and discussed the risks caused by large-scale service systems (Pang & Whitt, 2009). Meanwhile, capacity management translates into a complex process in modern call centers. Researchers have thus investigated the impacts of flexible labor resources (Kesavan et al., 2014) and attempted to develop real-time schedule adjustment frameworks (Mehrotra et al., 2010). Another prevalent technology-enabled change in call center operation is outsourcing; a wealth of research has explored issues related to outsourcing (Kocaga et al., 2015) and call-routing (Gans & Zhou, 2007) strategies in such contexts. In the above studies, researchers have mainly focused on optimizing system designs in contexts where technologies have been deployed to facilitate service operations (e.g., Aksin et al., 2007). However, little research has explored the effects of different technology-mediated contact designs with customers directly involved in service delivery (A. V. Roth & Mentor, 2003).

Specifically, in call center customer service, one typical technology-mediated service contact design is the IVR system, which enables self-service at the front end of phone calls (Tezcan & Behzad, 2012). Well-implemented IVR systems have the potential to automate a significant portion of services and lead to improved customer service experiences (Tezcan & Behzad, 2012). Thus, ample prior work has examined the design of IVR-equipped service systems (Khudyakov et al., 2010; Suhm & Peterson, 2002). Meanwhile, related studies from the user's perspective reveal that customers often feel frustrated when they interact with an IVR system because they perceive the services provided by IVR systems to be less customized and report that such systems occasionally do not understand their needs (Dean, 2008). Consequently, customers often attempt to avoid IVR systems due to the lack of personalized services or social interactions; instead, they seek direct interaction with human agents (Tezcan & Behzad, 2012).

Recent developments in AI technologies have enabled its applications in various contexts (Brynjolfsson et al., 2019; Cui et al., 2021; Sodhi et al., 2022). For example, in 2017, Google's ML algorithms achieved a 95% accuracy rate for speech recognition in the English language, a level that is close to actual human dialogue.[5] In the customer service setting, voice-based AI systems can understand customer needs through their voice inputs and can interact with customers in a human-like manner (Van Doorn et al., 2017; Wilson & Daugherty, 2018; Xiao & Kumar, 2021). However, considering the complexity of AI technology, it is challenging to determine the effectiveness of AI systems in a real-world setting that goes beyond training data (Brynjolfsson & Mcafee, 2017). Once an AI system is deployed, it is expected to handle a large variety of situations that may be unforeseen in training data. For example, customers may speak with accents or dialects while communicating with AI, thereby resulting in speech-recognition failures that influence the effectiveness of systems. Therefore, how AI implementation affects customer behaviors, and the performance of customer services remains an important empirical question that warrants further investigation.

3 | THE EFFECTS OF THE IMPLEMENTATION OF AI ON CALL CENTER CUSTOMER SERVICE

Based on the above discussions, in this section, we seek to discuss a few predictions on how the implementation of a voice-based AI system in call center customer service will affect three key metrics that are of interest to OM researchers: call length, demand for human service, and customer complaints. While we do not provide any directional hypotheses in this section, the discussion serves as a theoretical basis that guides our empirical analyses, which we report in subsequent sections.

3.1 | Call length

Call length represents the duration of a customer's service call (Gans et al., 2003), which is important in the management of call center customer service operations because it directly impacts scheduling and routing designs (Gans et al., 2003). In a traditional IVR system, the services are organized in a tree-like hierarchical structure, whereby the leaves represent different services, the nodes indicate customer states in the system, and the connections among different nodes indicate the paths to specific services. All paths are pre-designed and customers can move only from one node to another by inputting information in accordance with the guidance of the system. Typically, customers must pass through several nodes before reaching certain services. Meanwhile, they must pay attention to obtaining information on how to move from one node to another. An IVR system design typically entails a time-consuming service experience. In contrast, with an AI-based service system, customers can skip all the layers of IVR structures and directly access intended services by briefly summarizing their needs to the system, which is likely to result in shorter call lengths, compared to a traditional IVR system.

Conversely, it is also possible that an AI system leads to an increase in call lengths, as compared to IVR systems, due to the characteristics of the speech-based interaction mode. To begin with, when using the AI system, customers need to take time to summarize their needs in the form of dialogues for the AI system to predict the intended services. Second, according to prior research on communication modes, individuals interacting with text-based service systems (e.g., by inputting numbers in the IVR systems) follow the cognitive economy principle, such that they are more likely to focus on service requests and use keyword commands to improve communication efficiency (Le Bigot et al., 2007). In contrast, the speech-based interaction mode enhances users' involvement; users tend to use quest-irrelevant expressions, such as politeness expressions in their interactions (Chafe, 1982; Le Bigot et al., 2007), which could make information exchanges less effective (Le Bigot et al., 2007). In addition, when in conversational mode with an AI, users are expected to adapt their behavior to the interaction system (Cowan et al., 2015; Le Bigot et al., 2007). Consequently, users might devote more time and cognitive effort to formulate their speech and repeat information heard during the interactions in order to share a common lexicon and syntactic structure with the interaction system (Le Bigot et al., 2007). Therefore, speech-based AI service interactions may have longer service durations than services handled by an IVR system. Based on the above discussions, it is challenging to clearly predict the direction as well as the magnitude of changes in call length after the introduction of the voice-based AI system; thus, the effect of the voice-based AI system (vs. IVR system) remains an open question that warrants further empirical investigation.

3.2 | Demand for human service

Customers' demand for human service has direct implications for staffing problems and operational costs within call center customer service (Tezcan & Behzad, 2012). The introduction of AI may have mixed effects on customers' demand for human service. On the one hand, prior studies on AI applications demonstrate that, in certain contexts, individuals have a subjective perception against AI and, thus, might be reluctant to interact with it even though AI now offers high-level performance (Dietvorst et al., 2015; Longoni et al., 2019; Luo et al., 2019). When the AI system offers the flexibility of transferring to human agents, customers may skip interacting with AI and turn directly to human agents. Therefore, the AI-based system may increase customers' demand for human service.

On the other hand, previous research also suggests that providing individuals with even a slight amount of control over the AI's behavior has the potential to mitigate their aversion to it (Dietvorst et al., 2018), and this could be the case in our study. For example, the AI system enables customers to control the pace of service and customers have the freedom to decide when to transfer to human agents. Such a user-friendly design may mitigate customers' aversion to interacting with

AI systems as well as mitigate any potential increase in customers' demand for human service. Considering both sides of the arguments, it is unclear whether and to what extent the implementation of a voiced-based AI system would influence customers' demand for human services; thus, we seek to test this relationship empirically.

3.3 | Customer complaints

Customer complaints are manifestations of customers' negative service experience (Singh, 1988). Firms expend significant effort to improve customer service experience and reduce customer complaints. According to the service operations literature, firms create standardized service routines to control service delivery and ensure a uniform service level (Leidner, 1993). Standardized service routines reflect the preference of service providers with regard to the manner in which customer needs must be met, with the process steps being organized in a particular order. The service processes are largely determined by average customer demands and preferences (Victorino et al., 2013). Since they follow service routines designed for an average customer, service systems lack flexibility, cannot spontaneously react to unforeseen situations (Groth et al., 2009), and are likely to overlook customer heterogeneities (Ashforth & Fried, 1988). Dealing with customer heterogeneities (e.g., request and preference variations) was a major challenge for service operations (Frei, 2006), and flexibility is one of the important capabilities in operational design (Aksin et al., 2007; De Groote, 1994).

A diverse environment and heterogenous needs are best fitted with flexible technology (de Groote, 1994). Specifically, improving the degree of flexibility in how service systems react to customer requests enables the delivery of customized services (Tansik & Smith, 1991), thereby enabling an enhancement of customers' service experiences (S. Roth et al., 2006). For example, Victorino et al. (2013) reported that customers' perceived lower service quality from dinner recommendation services provided by an employee who rigidly follows the service script. In contrast, customers gave high ratings to service interactions in which the employee offers the flexibility of reacting to customer varieties (Victorino et al., 2013). In addition, Heim and Sinha (2002) showed that the flexibility of the service process in electronic retailing is positively associated with customer satisfaction. However, increased flexibility in a rule-based service system could potentially be accompanied by an increased complexity of the system, which makes it more likely to result in subjective service failures.

In the context of our study, the voice-based AI system accommodates customers' communication preference heterogeneities by enabling customers to express their service needs in ways that are most suitable for them. Meanwhile, compared with the IVR system, the AI system enhances the flexibility of service flows so that customers can directly locate their intended services, switch among different services, and transfer to human agents whenever they want. Therefore, we expect the implementation of the voice-based AI system to enhance customers' experience and reduce customer complaints; moreover, we also seek to empirically evaluate this effect.

4 | BACKGROUND AND DATA

Our study considers a NFE conducted by a large telecommunication company's call center customer service, which serves as an important channel for customer-firm interaction (Aksin et al., 2007). The company was established in 1995 and now has 14 branches and over 8000 employees, providing services to over three million customers in a major city (covering an area of approximately 53,100 km^2) in northeast China, with a market share of 33%. The company rolled out its voice-based AI system in its call center customer service system based on the last digit of customer phone numbers. Figure 1 summarizes the timeline of the NFE. Before December 19, 2018, all service calls were connected to the IVR system. From December 19, 2018, the AI system was implemented to replace the IVR system for a certain portion of the company's customers, which was chosen on the basis of a set of randomly selected last digits of the customer's phone number. Specifically, service calls from phone numbers with the last digit 1 or 7 were connected to AI, while calls from other phone numbers remained connected to the IVR system. Between December 19 and December 31, 2018, the updated service system was in the beta-testing phase and was not connected to the internal databases. Therefore, no service records were stored. Beginning on January 1, 2019, the AI system was connected to internal databases with phone records stored. Thereafter, beginning January 10, 2019, in addition to service calls from phone numbers ending in 1 or 7, service calls from phone numbers ending in 3, 5, or 9 were also connected to the AI system. After January 15, 2019, the AI system completely replaced the IVR system.

While interacting with the AI-based system, customers verbally state their requests briefly, and the AI system provides instant responses based on the analysis of information input by the customers. If customers do not describe their requests clearly, the AI system asks specific questions to guide customers to providing more information so that the AI can route them to the specific services in order to meet their needs (e.g., payments, check balance, temporarily stop service). If the IVR system or AI system is unable to provide specific services that customers are looking for, the customers have the option to be transferred to human agents. When interacting with the IVR system, customers need to strictly follow pre-designed service flows. After navigating the entire voice guidance on possible services, the system tells customers to press a specific number to be transferred to human agents. In contrast, the AI-based system sets no restrictions on when and how customers can transfer to human agents. At the beginning of the service, the AI system tells customers,

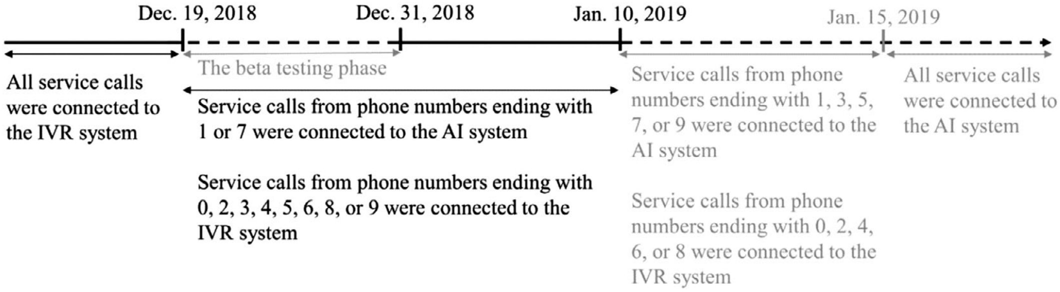

FIGURE 1 Timeline of the natural field experiment. AI, artificial intelligence; IVR, interactive voice response.

"If you need help from human agents, please say *'Transfer to human agents'*."

4.1 | Data and measures

The implementation of the voice-based AI system in call center customer service provides exogenous variations on the type of service system (voice-based AI vs. IVR system) that a customer experiences. The observation duration in our study was 30 days, including a 21-day pre-treatment period (November 28 to December 18, 2018) and a 9-day treatment period (January 1–9, 2019).[6] We exclude the beta-testing phase between December 19 and December 31, 2018, because we were unable to observe the outcome variables during this phase.

Our dataset contains timestamps of customers' phone call records, such as the start and end times of service calls, customers' profile information, such as age and gender and when a customer began using the telecommunication service, as well as the transcripts of customer–AI conversations from all the customers served by the company. We constructed the variable *Call Length* to measure a service call's duration using the difference between a call's start and end times. Considering the skewed distribution of *Call Length* (Gans et al., 2003), we log-transformed this variable in our estimations. In addition, we constructed the variable *Human Service* to capture whether a customer transferred to human agents in the service call. Further, extending the related research on customer service experiences in OM (Aksin et al., 2007), we considered customer complaints as one typical consequence of negative service experiences. Here, we constructed the variable *Customer Complaint* to capture whether the customer complained about the service within 30 min after a service call.

Following prior research on technology acceptance (Venkatesh et al., 2012), we also included a few additional variables on the observable individual characteristics to understand whether individual differences (e.g., age, gender, experience) moderate users' acceptance and use of the voice-based AI system. Specifically, we observed user age and gender. In addition, we constructed the variable *Service Tenure* to measure how many years a customer has been using his or her phone number and consider it a proxy for the customer's experience using the IVR system. Table 2 presents a summary of the operationalization of the main variables.

TABLE 2 Variables and definitions

Variables	Definitions
Call Length	The duration of a service call. It is measured by the difference between a call's end time and start time
Human Service	Whether a customer chose to transfer to human agents, with yes = 1 and no = 0
Customer Complaint	Whether a customer complained about the service within 30 min, with yes = 1 and no = 0
Age	The actual age of a customer calculated based on his/her birthday information
Gender	Female = 1 and male = 0
Service Tenure	The number of years a customer has been using his/her phone number. It is regarded as a proxy for the customer's experience using the IVR system

5 | METHODOLOGY AND RESULTS

5.1 | Econometric identification

First, we processed the data to ensure comparability between our treatment and control groups in the company's NFE. In the experiment, the group assignment hinges on the last digit of the customers' phone number, instead of randomly assigning customers to either the treatment or control group. To ensure comparability of observations, given the lack of individual random assignment, we tried to balance our samples before data analyses. If certain individuals prefer an even number as the last digit, it may result in significant differences between the two groups (i.e., even numbers vs. odd numbers); this is also observed in our data (see Supporting Information B). To address this issue, we first excluded data from customers with phone numbers that end in even numbers (i.e., 0, 2, 4, 6, and 8). Further, we conducted a series of pairwise comparisons of the observable covariates (i.e., *Age*, *Gender*,

TABLE 3 Comparisons of observable covariates and pre-treatment values of outcome variables

Last Digit	*Age*	*Gender*	*Service Tenure*	*Log (Call Length)*	*Human Service*	*Customer Complaint*
7	43.579 (11.516)	0.639 (0.480)	8.507 (3.740)	4.453 (0.953)	0.320 (0.466)	0.014 (0.116)
9	43.954 (11.590)	0.648 (0.478)	8.650 (3.728)	4.470 (0.922)	0.316 (0.465)	0.012 (0.109)
p-value	0.466	0.993	0.207	0.345	0.658	0.455

Note: Standard errors are given in parentheses. We observed insignificant differences in the observable covariates and the pre-treatment values of the outcome variables, thereby suggesting comparability of the groups with customers whose phone numbers end with 7 versus 9.

TABLE 4 Descriptive statistics

Variables	Observations	Mean	SD	Min	Max	Median
Log (Call Length)	18,580	4.446	0.869	2.996	7.365	4.407
Human Service	18,580	0.288	0.453	0	1	0
Customer Complaint	18,580	0.013	0.114	0	1	0
Age	18,580	43.765	11.554	15	70	43
Gender	18,580	0.643	0.479	0	1	1
Service Tenure	18,580	8.578	3.734	2	23	8

Service Tenure) and the pre-treatment values of outcome variables by the last digit of the customers' phone numbers (see Table B1 in Supporting Information B). We found no significant differences in the observable covariates across the groups of customers whose phone numbers end with 7 and 9 as indicated in Table 3. Table 4 reports the descriptive statistics of the main variables.

To provide additional model-free evidence for the treatment effect and comparability of the two groups, we also show the pre-treatment parallel trends of the outcome variables in Figures C1–C3 in Supporting Information C. Overall, the evidence indicates that customer groups with the last digit as 7 versus 9 are comparable. Thus, we opt to use the customer groups whose phone numbers ended with 7 (treated) and 9 (control) to estimate the effects of the AI-based system on the outcome variables in our main analyses.

Thereafter, we performed DID analyses to estimate the effects of the voice-based AI system on service call length, customers' need for human service, and customer complaints. We specify the DID estimations with Equations (1)–(3). In addition, we also incorporate customer-level random effects in these estimations.[7]

$$\begin{aligned} &Log(Call\ Length)_{it} \\ &\quad = \beta_0 + \beta_1 AI_{agent_i} + \beta_2 AI^{*}_{agent_i} After_{AI_t} \\ &\quad\quad + \beta_3 Age_i + \beta_4 Gender_i + \beta_5 Service\ Tenure_i \\ &\quad\quad + \mathrm{H} Day\ Dummy_t + u_i + \varepsilon_{it} \end{aligned} \tag{1}$$

$$Human\ Service\ Likelihood_{it} = \frac{Exp(U_{it})}{Exp(U_{it})+1};$$

$$\begin{aligned} U_{ist} &= \beta_0 + \beta_1 AI_{agent_i} + \beta_2 AI^{*}_{agent_i} After_{AI_t} + \beta_3 Age_i \\ &\quad + \beta_4 Gender_i + \beta_5 Service\ Tenure_i + \mathrm{H} Day\ Dummy_t \\ &\quad + u_i + \varepsilon_{it}. \end{aligned} \tag{2}$$

$$Customer\ Complain\ Likelihood_{it} = \frac{Exp(W_{it})}{Exp(W_{it})+1};$$

$$\begin{aligned} W_{ist} &= \beta_0 + \beta_1 AI_{agent_i} + \beta_2 AI^{*}_{agent_i} After_{AI_t} + \beta_3 Age_i \\ &\quad + \beta_4 Gender_i + \beta_5 Service\ Tenure_i + \mathrm{H} Day\ Dummy_t \\ &\quad + u_i + \varepsilon_{it}. \end{aligned} \tag{3}$$

In the equations above, i denotes customers, and t denotes observation time; *AI_agent* is a dummy variable, with 1 representing service calls from customers in the treatment group (i.e., customers whose phone numbers end with 7) and 0 representing service calls from customers in the control group (i.e., customers whose phone numbers end with 9). *After_AI* is a dummy variable that equals 1 for observations that took place after the introduction of the AI system and 0 for observations on or prior to December 18, 2018; $Day\ Dummy_t$ is a vector of time dummies representing each day during our observational period; u_i represents customer-specific random effects; and ε_{it} is the error term. In Equation (1), the dependent variable is *Log* (*Call Length*). In Equations (2) and (3), we observed the binary indicators of whether a call is transferred to human service and whether a call service eventually received a customer complaint; we estimated these outcomes with logistic regressions. We are interested in the coefficients

TABLE 5 Effects of AI implementation on call length

Variables	**(1) *Log* (*Machine_Call Length*)**	**(2) *Log* (*Machine_Call Length*)**	**(3) *Log* (*Human_Call Length*)**	**(4) *Log* (*Human_Call Length***	**(5) *Log* (*Call Length*)**	**(6) *Log* (*Call Length*)**
AI_agent	0.003 (0.018)		0.021 (0.071)		0.005 (0.022)	
*AI_agent * After_AI*	0.055** (0.021)	0.041* (0.022)	0.082 (0.082)	0.012 (0.083)	0.050** (0.023)	0.032 (0.023)
Age	−0.003*** (0.001)		−0.027** (0.003)		−0.006*** (0.001)	
Gender	−0.024 (0.017)		−0.022 (0.065)		−0.027 (0.021)	
Service Tenure	−0.004* (0.002)		−0.022** (0.009)		−0.006** (0.003)	
Observations	18,580	18,580	18,580	18,580	18,580	18,580
Between *R*-square	0.080	0.061	0.045	0.006	0.043	0.013
Number of customers	3625	3625	1818	5359	3625	3625
Day dummies	Y	Y	Y	Y	Y	Y
Customer random effects	Y	–	Y	–	Y	–
Customer fixed effects	–	Y	–	Y	–	Y

Note: Standard errors are given in parentheses. In columns 3 and 4, the values of *Human_Call Length* are 0 for services successfully handled by the AI system or IVR system, and we calculated *Log* (*Human_Call Length*) = *log* (*Human_Call Length* + 1).
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

TABLE 6 Effects of AI implementation on human service and customer complaint

Variables	**(1) *Human Service***	**(2) *Human Service***	**(3) *Customer Complaint***	**(4) *Customer Complaint***
AI_agent	0.007 (0.083)		0.129 (0.369)	
*AI_agent * After_AI*	0.103 (0.089)	0.036 (0.092)	−1.037** (0.406)	−1.101** (0.438)
Age	−0.035*** (0.004)		−0.069*** (0.017)	
Gender	−0.030 (0.079)		0.249 (0.358)	
Service Tenure	−0.025** (0.011)		0.117** (0.048)	
Observations	18,580	10,621	18,169	959
Number of customers	3625	1658	3625	107
Day dummies	Y	Y	Y	Y
Customer random effects	Y	–	Y	–
Customer fixed effects	–	Y	–	Y

Note: Standard errors are given in parentheses. ***$p < 0.01$, **$p < 0.05$, and *$p < 0.1$. In Columns 2 and 4, some observations were excluded when we conducted logistic regressions considering customer fixed effects.

of the interaction term, *AI_agent * After_AI*, as they capture the effects of the AI system (compared with the IVR system) on the outcomes. For example, if the coefficient in Equation (1) (i.e., β_2) is positive and statistically significant, it suggests that, compared with control customers—who used the IVR system and did not access the AI system—the treated customers, who did use the AI system experienced longer average call length after the implementation of the AI system.[8]

5.2 | Main findings

The regression estimations are presented in Tables 5 and 6, demonstrating the effects of the AI system on *Log* (*Call Length*), *Human Service,* and *Customer Complaint*. To explore how the durations of machine and human calls change after the implementation of the AI system, we separated the duration of machine service (*Machine_Call Length*) from human service (*Human_Call Length*)[9] and then conducted regressions. The results presented in columns 1−4 in Table 5 suggest that the implementation of voice-based AI significantly increases the duration of machine service by 5.65% (i.e., $100 \times (e^{0.055}{-}1)\%$) but exerts no effect on the duration of human service. These results indicate that customers tend to spend more time interacting with the AI system as compared to the IVR system. The results echo previous literature stating that, compared to mechanical self-service systems, customers tend to engage more with natural language-based service robots (Huang & Rust, 2021).

Specifically, diving into the records of customer-AI conversations, we found evidence for the wide use of quest-irrelevant characteristics (e.g., the use of first- or second-person pronouns, politeness expressions, and hesitation expressions), suggesting enhanced involvement during speech-based interactions, which can lead to an increase in the service duration (Hauptmann & Rudnicky, 1988; Le Bigot et al., 2007).[10] Meanwhile, per the results in columns 5 and 6 in Table 5, we found inconclusive evidence for the effects of the voice-based AI system on the total call length (i.e., *Log* (*Call Length*)).

Table 6 reports the results from estimating Equations (2) and (3). Specifically, the estimates presented in columns 1 and 2 in Table 6 suggest that the introduction of a voice-based AI system does not appear to exert a significant effect on customers' demand for human service, even though it is easier for customers to transfer to human agents when interacting with the voice-based AI system. The above results provide null evidence on the possible negative consequence of implementing voice-based AI in supporting customer service. When interacting with the AI system, customers can choose to transfer to human agents at the beginning of the services, which may increase the workload of human agents. However, we did not observe such an effect in our results. One possible explanation for the results is that the AI-based service system in our research context enables customers to control the service pace by transferring to human agents anytime they want. Giving customers the freedom to control AI can reduce their aversion against AI (Dietvorst et al., 2018).

We also considered the effects of AI implementation on *Customer Complaint*. In columns 3 and 4 in Table 6, we observed that the AI system significantly reduces customers' likelihood of filing complaint reports. We also quantified the magnitude of this effect in accordance with Hosmer et al. (2013). Specifically, compared with the average *Customer Complaint* before the implementation of AI in the sample ($M = 0.013$), we estimated that the implementation of AI reduces the probability of customer complaints to 0.005 (i.e., $0.013 * e^{-1.037}/(1 + 0.013 * e^{-1.037})$), with a decrease of 61.54% in customer complaints. As an extension to the literature that examines the impacts of AI-enabled features in OM, such as automation (Cui et al., 2021; Li & Li, 2022) and smartness (Cui et al., 2021), our results reveal that the service flexibility (a reflection of AI smartness) enhanced by the AI systems does, indeed, improve the overall service performance. In Supporting Information E, we replicated our main analysis and found similar results from the data from customers with phone numbers ending in odd numbers.

5.3 | Additional analyses

In the additional analyses, we first explored the heterogeneity in our main results (Section 5.3.1). Prior work has demonstrated that user characteristics play critical roles in affecting the performance of technology designs (Venkatesh & Morris, 2000; Venkatesh et al., 2012), and we thus tested the moderating effects of customer characteristics (e.g., age, gender, and service tenure). Next, we dived into the customer-AI conversations and considered the consequences of possible AI service failure (Section 5.3.2). Since AI cannot work perfectly to handle all service tasks, we tried to understand how customer-AI interaction and AI's speech-recognition failures influence customer service outcomes.

In addition, prior studies have found that individuals are more likely to accept and use new technologies when they get used to them (Taylor & Todd, 1995). In our research context, a customer can use the call service system several times during our observation period. Therefore, in Supporting Information F, we further analyzed whether the effects of AI implementation change as customers accumulate experience in interacting with the AI system, that is, the learning effects in customer-AI interactions. Our results suggest that, for relatively simple requests, the implementation of voice-based AI directly increases machine service duration and reduces customer complaints. When dealing with complex requests (service calls handled by human agents), the AI system only reduces customer complaints for customers who are experienced in using the AI system.

Finally, one potential explanation for the observed effects of the voice-based AI system on the outcomes of interests is the novelty effect. For example, customers may be unfamiliar with the AI system when the system is first introduced in the call center. In such a scenario, they are more likely to spend a longer amount of time interacting with the AI system or they may be more tolerant of the services provided by the AI system. In order to understand the possible novelty effect, in Supporting Information G, we re-estimated our regression equations in the main analysis by considering or eliminating records on the voice-based AI system's first- and second-time services for each customer. The results suggest that the implementation of the AI system persistently reduces customer complaints. However, possible novel effects of the AI system during the period of its early introduction indicate that the duration of machine service and customer demand for human service increases only temporarily after the introduction of the AI system and these effects are not significant in the long term.

5.3.1 | Heterogeneity by customer characteristics

After estimating the main effects of the AI system, we further examined how customer-level covariates—including age, gender, and experience of using the IVR system—moderate the effects of the voice-based AI system. The variable *Service Tenure* measures the number of years that a customer has been using his/her phone number, and we used it as a proxy for customers' experience using the IVR system. In terms of continuous variables, including *Age* and *Service Tenure*, we first mean-centered these variables before constructing the interaction terms. Table 7 presents the results of the moderation analyses.

As indicated in panel A of Table 7, we found that *Age* moderates the effects of the AI system. Specifically, older (vs. younger) customers may benefit more from the dialogue-based services supported by AI, such that after the implementation of voice-based AI system, they spend less time on call services, have less demand for human service, and register fewer complaints. In line with Meuter et al. (2005), who found that older customers are not proficient at using traditional IVR systems and thus are more reluctant to interact with these systems. Consequently, the convenience and flexibility enabled by the voice-based AI system are more helpful for improving the service experience of older customers. With regard to the moderating role of *Gender* in panel B, the AI-based system is more effective in reducing complaints from female customers (i.e., *Gender* = 0). Studies on the use of self-service technology indicated that females are strongly influenced by their perceptions of ease of use of technologies (Venkatesh & Morris, 2000). Therefore, they experience a significant improvement in AI-supported flexible services. Furthermore, we find that customers' experience of using the IVR system, *Service Tenure*, moderates the effect of the AI system on *Customer Complaint* (panel C). For customers who have more (vs. less) experience using the IVR system, the implementation of the AI system has a greater effect in reducing their complaints. One possible explanation for this is that experienced users are more familiar with the drawbacks of the IVR system and more likely to appreciate the benefits of the AI system, thereby tending to have fewer complaints. As an extension of prior literature that assumes service systems have the same service performance for all customers (Khudyakov et al., 2010; Tezcan & Behzad, 2012), our results indicate that the effects of AI-based systems vary in terms of customer gender, age, and service tenure.

5.3.2 | Speech-recognition failures in customer-AI interactions

We analyzed the transcripts of customer-AI conversations to examine how AI's speech recognition failures in customer-AI interactions may affect customers' demand for human service and customer complaints. To this end, we calculated the variable *Failure_Count* to measure the number of times that the AI system failed to recognize a customer's intention during a service call by counting the number of times the AI system repeated the same question. On average, during our observational window, approximately 28.5% of the customer-AI system service sessions involved speech-recognition failures. We also measured *Conversation_Count* to quantify the rounds of interaction between the AI system and a customer during a service call. Considering the skewed distributions of the variables *Failure_Count* and *Conversation_Count*, we used the log-transformed values of these variables in our regression estimations.[11] Table 8 presents the results pertaining to speech recognition failures of AI. The results suggest that *Log*(*Conversation_Count*) is negatively related to *Human Service* and *Customer Complaint*. One possible explanation is that, for service requests that can be handled by the AI system, customers tend to have more interaction rounds with the AI system and are less likely to turn to human agents and complain about the service. Meanwhile, *Log*(*Failure_Count*) is positively related to *Human Service* and *Customer Complaint,* thereby indicating that speech-recognition failures in customer-AI conversations can lead to significant and negative effects on call center service performance by increasing customers' demand for human service, thus leading to more customer complaints.

6 | DISCUSSION

6.1 | Key findings

This study investigates how the implementation of a voice-based AI system in call center customer services affects customer behavior and call center performance. The results reveal several interesting findings. First, we find that the voice-based AI system temporarily increases the duration of machine service and customers' demand for human service when the system is first introduced to customers, but these effects were not significant after the customers gained experience with the AI system. Second, the effects of the AI system on customer complaints vary in accordance with the complexity of the customers' service requests and the customers' experience of using the AI system. Specifically, the AI-system effectively reduces customer complaints for both experienced and inexperienced customers when customers have relatively simple requests. In contrast, with regard to complex requests, the AI system improves customers' service experience only after they accrue sufficient experience interacting with the AI system. Third, we explore how customer characteristics moderate the effects of the AI system. The results reveal that the AI system is more helpful in reducing complaints for older customers, female customers, and customers who are experienced in using the IVR system.

6.2 | Theoretical and practical implications

Our study contributes to the related literature on the application of AI systems and call center customer service operations. To begin with, this work extends the literature on AI applications by improving the understanding of the effects of AI systems in the customer service setting. Previous studies have either focused on deploying AI-based algorithms to support or optimize operational processes from the technical perspective (Senoner et al., 2021; Sun et al., 2019; Yang et al., 2022) or examined the effects of different AI-enabled

TABLE 7 Heterogeneity by customer characteristics

Panel A. ***Moderation by age***	**(1)** ***Log (Call Length)***	**(2)** ***Log (Call Length)***	**(3)** ***Human Service***	**(4)** ***Human Service***	**(5)** ***Customer Complaint***	**(6)** ***Customer Complaint***
AI_agent	0.004 (0.022)		0.009 (0.084)		0.332 (0.398)	
*AI_agent * After_AI*	0.053** (0.023)	0.035 (0.023)	0.084 (0.090)	0.003 (0.093)	−1.441*** (0.443)	−1.613*** (0.486)
*AI_agent * After_AI * Age*	−0.004* (0.002)	−0.004** (0.002)	−0.014* (0.008)	−0.022** (0.009)	−0.109*** (0.041)	−0.146*** (0.047)
Age	−0.009*** (0.001)		−0.036*** (0.005)		−0.104*** (0.027)	
*AI_agent * Age*	0.001 (0.002)		0.000 (0.007)		0.058 (0.036)	
*After_AI * Age*	0.008*** (0.001)	0.008*** (0.001)	0.008 (0.006)	0.010 (0.006)	0.055** (0.028)	0.060** (0.029)
Gender	−0.027 (0.021)		−0.031 (0.079)		0.250 (0.364)	
Service Tenure	−0.006** (0.003)		−0.025 (0.011)		0.119** (0.049)	
Observations	18,580	18,580	18,580	10,621	18,169	959
Number of customers	3625	3625	3625	1658	3625	107
Day dummies	Y	Y	Y	Y	Y	Y
Customer random effects	Y	–	Y	–	Y	–
Customer fixed effects	–	Y	–	Y	–	Y
Panel B. ***Moderation by Gender***	**(1)** ***Log (Call Length)***	**(2)** ***Log (Call Length)***	**(3)** ***Human Service***	**(4)** ***Human Service***	**(5)** ***Customer Complaint***	**(6)** ***Customer Complaint***
AI_agent	0.005 (0.037)		−0.028 (0.140)		0.657 (0.673)	
*AI_agent * After_AI*	0.038 (0.039)	0.015 (0.039)	0.053 (0.152)	−0.057 (0.156)	−2.669*** (0.904)	−2.873*** (0.990)
*AI_agent * After_AI * Gender*	0.017 (0.478)	0.025 (0.049)	0.075 (0.188)	0.143 (0.193)	2.118** (1.015)	2.306** (1.116)
Age	−0.006*** (0.001)		−0.035*** (0.004)		−0.069*** (0.017)	
Gender	−0.043 (0.033)		−0.054 (0.124)		0.599 (0.587)	
*AI_agent * Gender*	0.001 (0.046)		0.056 (0.174)		−0.738 (0.809)	
*After_AI * Gender*	0.031 (0.034)	0.025 (0.034)	−0.048 (0.133)	−0.070 (0.138)	−0.634 (0.584)	−0.700 (0.629)
Service Tenure	−0.006** (0.003)		−0.025** (0.011)		0.118** (0.049)	
Observations	18,580	18,580	18,580	10,621	18,169	959
Number of customers	3625	3625	3625	1658	3625	107
Day dummies	Y	Y	Y	Y	Y	Y
Customer random effects	Y	–	Y	–	Y	–
Customer fixed effects	–	Y	–	Y	–	Y
Panel C. ***Moderation by Service Tenure***	**(1)** ***Log (Call Length)***	**(2)** ***Log (Call Length)***	**(3)** ***Human Service***	**(4)** ***Human Service***	**(5)** ***Customer Complaint***	**(6)** ***Customer Complaint***
AI_agent	0.005 (0.022)		0.002 (0.083)		0.117 (0.402)	
*AI_agent * After_AI*	0.050** (0.023)	0.032 (0.023)	0.111 (0.090)	−0.049 (0.092)	−0.991** (0.423)	0.998** (0.445)
*AI_agent * After_AI * Tenure*	−0.007 (0.006)	−0.006 (0.006)	0.033 (0.025)	0.036 (0.025)	−0.257** (0.110)	−0.217* (0.116)
Age	−0.006*** (0.001)		−0.035*** (0.004)		−0.073*** (0.019)	
Gender	−0.027 (0.021)		−0.030 (0.079)		0.246 (0.390)	
Service Tenure	−0.015*** (0.004)		−0.037** (0.017)		0.023 (0.082)	
*AI_agent * Service Tenure*	0.002 (0.006)		−0.010 (0.023)		0.124 (0.108)	
*After_AI * Service Tenure*	0.021*** (0.004)	0.020*** (0.004)	0.026 (0.018)	0.025 (0.018)	0.198*** (0.075)	0.170** (0.078)
Observations	18,580	18,580	18,580	10,621	18,169	959
Number of customers	3625	3625	3625	1658	3625	107
Day dummies	Y	Y	Y	Y	Y	Y
Customer random effects	Y	–	Y	–	Y	–
Customer fixed effects	–	Y	–	Y	–	Y

Note: Standard errors are given in parentheses. In columns 4 and 6, some observations were excluded when we conducted Logistic regressions considering customer fixed effects. ***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

TABLE 8 Effects of details in customer–AI interactions

Variables	(1) *Human Service*	(2) *Human Service*	(3) *Customer Complaint*	(4) *Customer Complaint*
Log (Conversation_Count)	−3.858*** (0.096)	−1.971*** (0.104)	−3.347*** (0.608)	−1.122** (0.461)
Log (Failure_Count)	0.966*** (0.090)	0.242* (0.128)	1.610*** (0.594)	0.213 (0.585)
Age	−0.033*** (0.003)		−0.027 (0.017)	
Gender	−0.122* (0.067)		0.136 (0.388)	
Service Tenure	−0.042*** (0.009)		−0.057 (0.052)	
Observations	17,274	6950	17,274	366
Number of customers	9042	1880	9042	78
Day dummies	Y	Y	Y	Y
Customer random effects	Y	–	Y	–
Customer fixed effects	–	Y	–	Y

Note: Standard errors are given in parentheses. In columns 2 and 4, most observations were excluded when we conducted logistic regressions considering customer fixed effects. In column 4, the coefficient of *Log* (*Failure_Count*) is not significant and one possible reason for the result may be that some observations were excluded when we conducted logistic regressions considering customer fixed effects.

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

features in contexts such as price request (Cui et al., 2021), order decision-making (Li & Li, 2022), and automated pricing (Karlinsky-Shichor & Netzer, 2019). These studies mainly investigated certain advantages and drawbacks of AI-enabled automation (Cui et al., 2021; Karlinsky-Shichor & Netzer, 2019; Li & Li, 2022) or smartness (Cui et al., 2021). As an extension, our study explores the effectiveness of AI-enabled service flexibility—a specific reflection of AI smartness—in call center customer service. Technology-based self-service systems (e.g., Automated teller machines (ATMs)) have already been demonstrated to work well in dealing with highly structured service tasks (Barua et al., 1991), while our study suggests that AI-enabled service flexibility is more likely to improve the service effectiveness when dealing with tasks with high variability (e.g., call center services). Thus, Information Technology (IT) investment decisions in service operations should match the features of both service tasks and technologies.

In addition, prior studies have indicated that the realization of AI's value in the context of human-AI interaction crucially depends on institutional settings and the role that customers play in using AI applications (Dietvorst et al., 2018; Luo et al., 2019). Often, customers are reluctant to interact with AI when they passively receive AI-supported marketing information (Luo et al., 2019), forecasting results (Dietvorst et al., 2015), or medical care (Longoni et al., 2019). However, our study suggests that when customers have the freedom to control the flow and direction of the service, the AI-based service system does not result in a significant increase in demand for human service, even though the AI system allows customers to transfer to human agents at any time during the interaction. Furthermore, our results offer preliminary insight into the negative effects of speech-recognition failures on customer-AI system interactions, thereby enriching research on imperfect AI (Dietvorst et al., 2015, 2018).

Our study also contributes to the literature on call center customer service operations by investigating how AI technologies impact customer behavior and the performance of call center customer service. Previous research has examined changes in call center customer service operations elicited by technological advances, such as call center centralization (Adria & Chowdhury, 2004), flexible resource management (Kesavan et al., 2014), and outsourcing (Kocaga et al., 2015), but limited attention has been paid to exploring the effects of the contact designs of different technology-mediated services with direct customer involvement (Froehle & Roth, 2004; A. V. Roth & Mentor, 2003). Moreover, the OM literature mainly focuses on measuring the performance of call center customer services from the firm's perspective, using easily trackable metrics, such as operational costs (Tezcan & Behzad, 2012) and wait time (Khudyakov et al., 2010; Singhal et al., 2019). There is limited research on customers' service experience (Aksin et al., 2007). Through the customers' perspective, our study examines the effects of an AI-based service system on customer complaints, which is a key consequence of customers' negative service experiences. We find that customers tend to make fewer complaints when served by the voice-based AI system. In addition, enriching prior studies that treat customers as homogeneous and use a single metric to represent the performance of service systems for all customers (Khudyakov et al., 2010; Tezcan & Behzad, 2012), this study further examines how the effects of the AI-based system vary in accordance with customers characteristics, such as age, gender, and experience in using the traditional IVR system.

Furthermore, our findings also have important practical implications. First, we find that the implementation of the voice-based AI system in call center customer services helps improve customer service experiences (i.e., reduces customer complaints) and that the flexibility of transferring to human agents, enabled by the AI system, does not lead to a significant increase in customers' demand for human service in the long term. These findings showcase the value of voice-based AI systems in the provision of customer service.

Thus, companies can continue implementing AI systems to support customer services. Second, our findings also shed light on bridging the proof-of-concept-to-production gap (Perry, 2021). We find that the effectiveness of the AI system is closely dependent on the service tasks (e.g., the complexity of customers' service requests) and customers' experience of using AI systems. As predicted, the implementation of the AI system directly improves customers' service experience in relatively simple service tasks. In terms of handling complex requests, the voice-based AI system only operates effectively to reduce complaints from customers who have gained enough knowledge about the interacted AI system. Thus, users may suffer the proof-of-concept-to-production gap in the early stages of adoption, particularly when dealing with complex tasks (Sodhi et al., 2022). Correspondingly, customer service operations that are equipped with AI systems can initially distinguish simple customer requests from complex ones based on historical service records and then encourage customers with simple requests to use AI-assisted services. Platforms can consider guiding customers to establish appropriate expectations of the AI system and transfer customers with complex requests to human agents as quickly as possible. Third, our results indicate the possible novelty effect of a newly implemented AI system. We find that the duration of machine service and customers' need for human service temporally increases after the introduction of the AI system. Companies can take these effects into account when scheduling resources for their newly implemented AI-based service systems. Moreover, in our study, suggestive evidence from customer-AI conversations reveals that customers are more likely to turn to human agents and complain about services after experiencing speech-recognition failures. This is likely not a huge concern in the longer run, as AI-based services are likely to improve in terms of speech recognition, given their learning capabilities.

6.3 | Limitations and future research

Our study has several limitations, which also indicate ample opportunities for future research. First, our experimental randomization is based on the last digit of customers' phone numbers rather than being performed at the individual level, and thus we balanced our samples before data analysis. Second, due to data limitations, we could not observe detailed records of specific service requests handled by the IVR system; therefore, we were unable to categorize service requests based on objective service types. It will be interesting for future researchers to extend our findings based on the objective complexity of customer service requests. Third, the current study focuses on the effects of a voice-based AI system on call length, customers' demand for human service, and customer complaints. Future research could explore other outcome variables, such as service satisfaction, customer retention, and future customer engagement, which reflect the value of AI implementation for businesses. Moreover, leveraging the transcripts of customer-AI conversations, we conducted a few preliminary analyses on customer-AI interactions by examining the negative effects of speech-recognition failures. It would be interesting for future research to explore other factors in human-AI interactions, such as emotions expressed by AI and service tones used by AI in conversations, which inform the design of voice-based AI systems. Last, we analyze the effectiveness of deploying a voice-based AI system to replace the traditional IVR system in the telecommunication customer service setting in China. The generalizability of our findings might be subject to the technical designs of the AI and IVR systems and cultural variation, as well as differences in levels of technology development among countries, all of which may affect users' attitude to and adoption of AI. Thus, we encourage future research to further explore the implications of voice-based AI systems among different user populations or in other service settings.

ORCID

Ni Huang https://orcid.org/0000-0003-3416-513X
Yili Hong https://orcid.org/0000-0002-0577-7877
Xunhua Guo https://orcid.org/0000-0002-7243-4292

ENDNOTES

[1] For example, "Google's Duplex Uses A.I. to Mimic Humans (Sometimes)," *New York Times*. https://www.nytimes.com/2019/05/22/technology/personaltech/ai-google-duplex.html.

[2] Forbes Insights, "How AI Is Revamping the Call Center." https://www.forbes.com/sites/insights-ibmai/2020/06/25/how-ai-is-revamping-the-call-center/?sh = 1f42cf1034b2.

[3] Mordor Intelligence, "AI Market in Call Center Applications—Growth, Trends, Covid-19 Impact, and Forecasts (2021–2026)." https://www.mordorintelligence.com/industry-reports/ai-market-in-call-center-applications.

[4] An NFE is the type of experiment "where the environment is one where the subjects naturally undertake these tasks and where the subjects do not know that they are participants in an experiment. Such an exercise represents an approach that combines the most attractive elements of the lab and naturally-occurring data: randomization and realism" (List, 2007).

[5] Google's ability to understand natural language is almost equivalent to that of humans. https://www.vox.com/2017/5/31/15720118/google-understand-language-speech-equivalent-humans-code-conference-mary-meeker.

[6] In this NFE, we have a 9-day treatment period (between January 1 and January 9, 2019, when the customers with the last digit of phone number 7 connected to AI system versus those with the last digit of phone number 9 connected to IVR system). Also, a 3-week pre-treatment period allows us to check the parallel trend of our key outcome variables before the experiment (from November 28 to December 18, 2018, when all service calls were connected to the IVR system). In addition, the cooperating telecommunication company designs its services on a monthly basis (e.g., customers choose a monthly service package, charge bills for the next month). Taken together, we chose a 30-day observation window.

[7] In this paper, we selected random-effects models for two reasons. First, in our field experiment, randomization occurred based on the last digit of the phone number. Thus, for the same customer, all his/her service records were either in the treatment group or the control group. When conducting regressions with customer fixed effects, a few variables would have been subsumed, like the variable *AI_agent* that indexes whether a record was in the treatment or control group, and dummies that capture the features of specific customers (e.g., *Age*, *Gender*, *Service Tenure*) in the regression models. Second, for dummy variables *Human Service* and *Customer Complaint*, the records of these variables with a value of 1 are relatively

sparse. If we chose fixed-effects models, many observation groups would have been omitted because of all negative outcomes (i.e., *Human Service* = 0 or *Customer Complaint* = 0). In order to check the robustness of our results, we also present the estimation results of fixed-effects models.

[8] We also conducted a series of placebo tests to check whether the identified effects existed before the introduction of the AI system or between customer groups who did not get access to the AI system. The results of the placebo tests are reported in Supporting Information D.

[9] *Machine_Call Length* measures the duration of a machine (i.e., AI or IVR) service, while *Human_Call Length* captures the duration of a service delivered by human agents.

[10] We examined the records of customer-AI conversations to provide possible explanations for the increase in call length (particularly in machine call length) and found that in 28.8% of the conversations, customers use first- or second-person pronouns (e.g., "I," "you"), suggesting enhanced involvement during conversations. In addition, approximately 21.3% of the conversations include at least one utterance to express politeness (e.g., "thank you"), and 16.0% of the conversations contain hesitation expressions (e.g., "Uh") when customers form utterances during conversations. Consistent with research on interaction mode, all these quest-irrelevant characteristics in speech-based interactions—although making the interaction more natural—can also lead to an increase in the service duration (Hauptmann & Rudnicky 1988; Le Bigot et al., 2007).

[11] In cases when variables include 0, we added 1 before the logarithm transformation. For example, we calculated *Log* (*Failure_Count*) = *log* (*Failure_Count*+1).

ACKNOWLEDGMENTS

The authors thank the department editor, Subodha Kumar, the anonymous senior editor, and reviewers for a most insightful and constructive review process. The manuscript has also benefited from the discussions with and feedback of the participants at the International Conference on Information Systems (ICIS). Lingli Wang acknowledges support from the Beijing University of Posts and Telecommunications Basic Scientific Research Program (2022RC21) and National Natural Science Foundation of China (72201038); Luning Liu acknowledges support from the National Natural Science Foundation of China (72034001, 71974044); Xunhua Guo acknowledges support from the Tsinghua University Initiative Scientific Research Program (2019THZWYX08) and the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (17JJD630006).

REFERENCES

Adria, M., & Chowdhury, S. D. (2004). Centralization as a design consideration for the management of call centers. *Information & Management*, *41*, 497–507.

Aksin, Z., Armony, M., & Mehrotra, V. (2007). The modern call center: A multi-disciplinary perspective on operations management research. *Production and Operations Management*, *16*(6), 665–688. https://doi.org/10.1111/j.1937-5956.2007.tb00288.x

Ashforth, B. E., & Fried, Y. (1988). The mindlessness of organizational behaviors. *Human Relations*, *41*(4), 305–329. https://doi.org/10.1177/001872678804100403

Barua, A., Kriebel, C. H., & Mukhopadhyay, T. (1991). An economic analysis of strategic information technology investments. *MIS Quarterly*, *15*(3), 313–331. https://doi.org/10.2307/249643

Beckman, S., & Sinha, K. K. (2005). Conducting academic research with an industry focus: Production and operations management in the high tech industry. *Production and Operations Management*, *14*(2), 115–124. https://doi.org/10.1111/j.1937-5956.2005.tb00013.x

Bitner, M. J. (1992). Servicescapes: The impact of physical surroundings on customers and employees. *Journal of Marketing*, *56*(2), 57–71. https://doi.org/10.1177/002224299205600205

Brynjolfsson, E., & McAfee, A. (2017). *The business of artificial intelligence*. https://hbr.org/2017/07/the-business-of-artificial-intelligence

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449–5460. https://doi.org/10.1287/mnsc.2019.3388

Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In Tannen, D. (Ed.), *Spoken and written language* (pp. 35–54). Ablex Publishing Corporation,.

Cowan, B. R., Branigan, H. P., Obregon, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling, and alignment effects on syntactic choices in human-computer dialogue. *International Journal of Human-Computer Studies*, *83*, 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, *27*(10), 1749–1769. https://doi.org/10.1111/poms.12707

Cui, R., Li, M., & Zhang, S. (2021). AI and procurement. *Manufacturing & Service Operations Management*, *24*(1), 83–97.

Dean, D. H. (2008). What's wrong with IVR self-service. *Managing Service Quality*, *18*(6), 594–609. https://doi.org/10.1108/09604520810920086

De Groote, X. (1994). The flexibility of production processes: A general framework. *Management Science*, *40*(7), 933–945. https://doi.org/10.1287/mnsc.40.7.933

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Fountain, T., McCarthy, B., & Saleh, T. (2019). Building the AI-powered organization. *Harvard Business Review*, *97*(4), 62–73.

Frei, F. X. (2006). Breaking the trade-off between efficiency and service. *Harvard Business Review*, *84*(11), 93–101.

Froehle, C. M., & Roth, A. V. (2004). New measurement scales for evaluating perceptions of the technology-mediated customer service experience. *Journal of Operations Management*, *22*, 1–21. https://doi.org/10.1016/j.jom.2003.12.004

Gans, N., Koole, G., & Mandelbaum, A. (2003). Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, *5*(2), 79–141.

Gans, N., & Zhou, Y. (2007). Call-routing schemes for call-center outsourcing. *Manufacturing & Service Operations Management*, *9*(1), 33–50.

Groth, M., Hennig-Thurau, T., & Walsh, G. (2009). Customer reactions to emotional labor: The roles of employee acting strategies and customer detection accuracy. *Academy of Management Journal*, *52*(5), 958–974. https://doi.org/10.5465/amj.2009.44634116

Hauptmann, A. G., & Rudnicky, A. I. (1988). Talking to computers: An empirical investigation. *International Journal of Man-Machine Communication*, *28*(6), 583–604.

Heim, G. R., & Sinha, K. K. (2002). Service process configurations in electronic retailing: A taxonomic analysis of electronic food retailers. *Production and Operations Management*, *11*(1), 54–74. https://doi.org/10.1111/j.1937-5956.2002.tb00184.x

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

Huang, M., & Rust, R. T. (2021). Engaged to a robot? The role of AI in service. *Journal of Service Research*, *24*(1), 30–41. https://doi.org/10.1177/1094670520902266

Karlinsky-Shichor, Y., & Netzer, O. (2019). *Automating the B2B salesperson pricing decisions: Can machines replace humans and when*. (Working paper), SSRN.

Kellogg, D. L., & Chase, R. B. (1995). Constructing an empirically derived measure for customer contact. *Management Science*, *41*(11), 1734–1749. https://doi.org/10.1287/mnsc.41.11.1734

Kesavan, S., Staats, B. R., & Gilland, W. (2014). Volume flexibility in services: The costs and benefits of flexible labor resources. *Management Science*, *60*(8), 1884–1906. https://doi.org/10.1287/mnsc.2013.1844

Khudyakov, P., Feigin, P. D., & Mandelbaum, A. (2010). Designing a call center with an IVR (interactive voice response). *Queueing Systems*, *66*, 215–237. https://doi.org/10.1007/s11134-010-9193-y

Kocaga, Y. L., Armony, M., & Ward, A. R. (2015). Staffing call centers with uncertain arrival rates and co-sourcing. *Production and Operations Management*, *24*(7), 1101–1117. https://doi.org/10.1111/poms.12332

Krishnan, M. S., Ramaswamy, V., Meyer, M. C., & Damien, P. (1999). Customer satisfaction for financial services: The role of products, services, and information technology. *Management Science*, *45*(9), 1194–1209. https://doi.org/10.1287/mnsc.45.9.1194

Kumar, S., Mookerjee, V., & Shubham, A. (2018). Research in operations management and information systems interface. *Production and Operations Management*, *27*(11), 1893–1905. https://doi.org/10.1111/poms.12961

Le Bigot, L., Terrier, P., Amiel, V., Poulain, G., Jamet, E., & Rouet, J. (2007). Effect of modality on collaboration with a dialogue system. *International Journal of Human-Computer Studies*, *65*(12), 983–991. https://doi.org/10.1016/j.ijhcs.2007.07.002

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Leidner, R. (1993). *Fast food, fast talk: Service work and the routinization of everyday life*. University of California Press.

Li, M., & Li, T. (2022). AI automation and retailer regret in supply chains. *Production and Operations Management*, *31*(1), 1059–1478. https://doi.org/10.1111/poms.13498

List, J. A. (2007). *Field experiments: A bridge between lab and naturally occurring data*. (Working paper), NBER.

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629–650. https://doi.org/10.1093/jcr/ucz013

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937–947.

Markets and Markets. (2021). *Conversational AI market report*. https://www.marketsandmarkets.com/Market-Reports/conversational-ai-market-49043506.html

Mehrotra, V., Ozluk, O., & Saltzman, R. (2010). Intelligent procedures for intra-day updating of call center agent schedules. *Production and Operations Management*, *19*(3), 353–367. https://doi.org/10.1111/j.1937-5956.2009.01097.x

Meuter, M. L., Biter, M. J., Ostrom, A. L., & Brown, S. W. (2005). Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. *Journal of Marketing*, *69*(2), 61–83. https://doi.org/10.1509/jmkg.69.2.61.60759

Pang, G., & Whitt, W. (2009). Service interruptions in large-scale service systems. *Management Science*, *55*(9), 1499–1512. https://doi.org/10.1287/mnsc.1090.1038

Perry, T. S. (2021). *AI pioneer says machine learning may work on test sets, but that's a long way from real world use*. https://spectrum.ieee.org/andrew-ng-xrays-the-ai-hype

Resnick, P., & Virzi, R. A. (1995). Relief from the audio interface blues: Expanding the spectrum of menu, list, and form styles. *ACM Transactions on Computer-Human Interaction*, *2*(2), 145–176. https://doi.org/10.1145/210181.210183

Roth, A. V., & Mentor, L. J. (2003). Insights into service operations management: A research agenda. *Production and Operations Management*, *12*(2), 145–164. https://doi.org/10.1111/j.1937-5956.2003.tb00498.x

Roth, S., Woratschek, H., & Pastowski, S. (2006). Negotiating prices for customized services. *Journal of Service Research*, *8*(4), 316–329. https://doi.org/10.1177/1094670506286330

Senoner, J., Netland, T., & Feuerriegel, S. (2021). Using explainable artificial intelligence to improve process quality: Evidence from semiconductor manufacturing. *Management Science*, *68*(8), 5704–5723.

Singh, J. (1988). Consumer complaint intentions and behavior: Definitional and taxonomical issues. *Journal of Marketing*, *52*(1), 93–107. https://doi.org/10.1177/002224298805200108

Singhal, K., Singhal, J., & Kumar, S. (2019). The value of the customer's waiting time for general queues. *Decision Sciences*, *50*(3), 567–581. https://doi.org/10.1111/deci.12343

Sodhi, M. S., Seyedghorban, Z., Tahernejad, H., & Samson, D. (2022). Why emerging supply chain technologies initially disappoint: Blockchain, IoT, and AI. *Production and Operations Management*, *31*(6), 2517–2537. https://doi.org/10.1111/poms.13694

Soteriou, A. C., & Chase, R. B. (1998). Linking the customer contact model to service quality. *Journal of Operations Management*, *16*(4), 495–508. https://doi.org/10.1016/S0272-6963(98)00026-6

Suhm, B., Bers, J., McCarthy, D., Freeman, B., Getty, D., Godfrey, K., & Peterson, P. (2002). A comparative study of speech in the call center: Natural language call routing vs. touch-tone menus. *Proceedings of the SIGCHI Conference on Human Factors Computing Systems*. Minneapolis, MN (pp. 283–290).

Suhm, B., & Peterson, P. (2002). A data-driven methodology for evaluating and optimizing call center IVRs. *International Journal of Speech Technology*, *5*, 23–37. https://doi.org/10.1023/A:1013674413897

Sun, C., Shi, Z., Liu, X., Ghose, A., Li, X., & Xiong, F. (2019). *The effect of voice AI on consumer purchase and search behavior*. (Working paper), SSRN.

Tang, M., Pellom, B., & Hacioglu, K. (2003). Call-type classification and unsupervised training for the call center domain. *IEEE Workshop on Automatic Speech Recognition and Understanding*, St. Thomas, Virgin Islands (pp. 204–208).

Tansik, D. A., & Smith, W. L. (1991). Dimensions of job scripting in services organizations. *International Journal of Service Industry Management*, *2*(1), 35–49. https://doi.org/10.1108/09564239110000127

Taylor, S., & Todd, P. (1995). Assessing IT usage: The role of prior experience. *MIS Quarterly*, *19*(4), 561–570. https://doi.org/10.2307/249633

Tezcan, T., & Behzad, B. (2012). Robust design and control of call centers with flexible interactive voice response systems. *Manufacturing and Service Operations Management*, *14*(3), 386–401. https://doi.org/10.1287/msom.1120.0378

Valuates Reports. (2022). *Call center AI market size to reach USD 9,949.61 million by 2030 at CAGR 26.3%*. https://www.prnewswire.com/news-releases/call-center-ai-market-size-to-reach-usd-9-949-61-million-by-2030-at-cagr-26-3—valuates-reports-301461943.html

Van Doorn, J., Mende, M., Noble, S. M., Hulland, J., Ostrom, A. L., Grewal, D., & Petersen, J. A. (2017). Domo arigato Mr. Roboto: Emergence of automated social presence in organizational frontlines and customers' service experiences. *Journal of Service Research*, *20*(1), 43–58. https://doi.org/10.1177/1094670516679272

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly*, *24*(1), 115–139. https://doi.org/10.2307/3250981

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Victorino, L., Verma, R., & Wardell, D. G. (2013). Script usage in standardized and customized service encounters: Implications for perceived service quality. *Production and Operations Management*, *22*(3), 518–534. https://doi.org/10.1111/j.1937-5956.2012.01382.x

Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, *96*(4), 115–123.

Xiao, L., & Kumar, V. (2021). Robotics for customer service: A useful complement or an ultimate substitute? *Journal of Service Research*, *24*(1), 9–29. https://doi.org/10.1177/1094670519878881

Yang, C., Feng, Y., & Whinston, A. (2022). Dynamic pricing and information disclosure for fresh produce: An artificial intelligence approach. *Production and Operations Management*, *31*(1), 155–171. https://doi.org/10.1111/poms.13525

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wang, L., Huang, N., Hong, Y., Liu, L., Guo, X., & Chen, G. (2023). Voice-based AI in call center customer service: A natural field experiment. *Production and Operations Management*, 1–17. https://doi.org/10.1111/poms.13953