

The Effect of Voice AI on Digital Commerce

Chenshuo Sun (csun@stern.nyu.edu)

Zijun Shi, HKUST

Xiao Liu, NYU Stern

Anindya Ghose, NYU Stern

This version: April 30, 2024

Abstract

Voice-activated shopping assistants (voice AI), such as Amazon Alexa or Alibaba Tmall Genie, have been gaining popularity worldwide as a new channel for online shopping. In this paper, we use large-scale archival data of consumer-level purchase records from Alibaba, the world's largest e-commerce platform, to empirically investigate how consumers' adoption of Tmall Genie affects their consumption. The results show that the average consumer's weekly spending on Alibaba increased by 16.6% within the first four months after adopting voice AI. Additionally, we explore specific product features that moderate the effect of Genie adoption by examining the repeat purchase, product substitutability and familiarity, supporting a mechanism that involves reducing information acquisition costs. The positive effects of Genie adoption remain significant on repeat purchase in the long term although they attenuate over time. Furthermore, our analyses reveal that on average, the voice channel has a positive spillover effect on spending on the PC channel but no significant effect on the mobile channel. The channel dynamics are contingent on specific shopping contexts. Our results demonstrate that voice AI devices with shopping capabilities can enhance the growth of the affiliated e-commerce platform. As the first study to empirically examine the impact of voice AI adoption on e-commerce consumption, our paper provides valuable implications for e-commerce platforms and retailers leveraging voice-activated shopping.

Keywords

Voice AI ▪ voice shopping ▪ e-commerce ▪ causal inference
▪ heterogeneity ▪ channel dynamics ▪ transactional cost

1. Introduction

In recent times, generative AI, notably exemplified by ChatGPT, has garnered pronounced attention for its remarkable capabilities, reflecting transformative potentials in creating humanlike texts and information (Li et al., 2024; Goli and Singh 2024; Brynjolfsson et al. 2023). The expanding utilization of conversational AI in market research and e-commerce facilitates a nuanced understanding of consumer preferences and scaling the AI-enabled e-commerce market. This market is expected to reach US \$16.8 billion by 2030, underscoring its pivotal role in deciphering and potentially, reshaping consumer behaviors and market research (Li et al., 2024; Forbes, 2023). Voice-activated smart devices, backed by conversational AI, are rapidly gaining popularity among companies and consumers. In the US, Amazon Alexa, Google Home, and Apple Homepod have become household names, while in China, Tmall Genie, Baidu Xiaodu, and Microsoft Xiaoice are widely adopted.¹

Coupled with the rise of these devices is the burgeoning trend of voice-activated shopping, where consumers interact with a voice AI interface to place orders. For instance, shopping through Amazon Alexa saw a threefold increase between the 2017 and 2018 holiday seasons. Walmart, recognizing this trend, partnered with Google in April 2019 to launch Walmart Voice Order, enabling Google Home users to order directly from Walmart's online portal using voice commands (Cohen, 2019). Surveys confirm this shift: Adobe Analytics reports that 22% of home speaker owners shop using their devices,² and PwC notes that 10% of US voice AI users regularly place voice-based orders.³

¹ Major features and uses of four popular voice assistant devices are provided in Appendix A1.

² Cited from <https://www.marketwatch.com/story/the-booming-smart-speaker-market-and-the-services-it-will-help-2018-01-18>.

³ Cited from https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html?utm_campaign=sbpwc&utm_medium=site&utm_source=articletext.

Despite the significance of voice shopping in consumers' daily lives and the digital economy, little is known about how the adoption of a voice-activated shopping device impacts a consumer's subsequent shopping behaviors. Voice shopping could plausibly affect a consumer's purchase journey. For example, if a consumer is browsing videos from a popular cosmetics blogger and encounters a video about lipsticks, the consumer can immediately start the search process by asking, "*Alexa, what are the best-selling lipsticks?*" If Alexa generates a desirable suggestion, the consumer can place an order in seconds by saying, "*Add to my cart and Check out.*"⁴ In this scenario, consumers start with an "I want-to-know moment" conversation with voice AI, followed by an "I want-to-buy moment" which can be accomplished with the help of voice AI.

The potential impact of voice AI device adoption on consumer spending remains ambiguous and warrants rigorous empirical examination. Two contradictory effects could be at work simultaneously. On one hand, voice AI devices enable consumers to shop hands-free⁵; the vocal search process can make it convenient for consumers to efficiently locate specific products while multitasking, potentially leading to increased spending⁶. Additionally, the ease of voice shopping might result in more impulsive purchases during interactions with the voice AI⁷. On the other hand, research on auditory consumer behavior indicates that information presented by voice could be more difficult to process than information presented in text or visually (Munz and Morwitz, 2022). Compounding this, inaccuracies in voice AI recognition not only cause inconvenience but also result in significant time losses, as consumers are often forced to restart the search process with increased forethought (Melumad, 2023). This increased cognitive effort required for spoken communication could discourage thorough product exploration and,

⁴ Cited from <https://www.racked.com/video/2018/2/20/17031958/alexa-shopping-amazon-echo-dot-voice>.

⁵ Cited from Walmart: <https://www.walmart.com/cp/voice-shopping/5028404>

⁶ Cited from <https://debutify.com/blog/voice-shopping/>

⁷ Cited from <https://voicebot.ai/2017/08/02/voice-meets-impulse-buy-thought-transaction-22-6-seconds/>

subsequently, reduce purchasing intentions. These opposing predictions create a tension that underscores the need for a comprehensive empirical study that identifies the causal effects of voice AI adoption on e-commerce consumption. Our research is thus crucial not only for advancing academic understanding but also for guiding e-commerce platforms in strategic decision-making.

To the best of our knowledge, our study is the first empirical examination of this topic and, therefore, this paper contributes to the growing field of research on the impact of emerging technologies on consumption (Xu et al., 2023; Adamopoulos et al., 2020; Pamuru 2021; Ghose, 2018) and studies on voice AI and chatbots (Melumad, 2023; Bergner et al., 2023; Son et al., 2023; Smith, 2020; Ba et al., 2020; Luo et al., 2019). Beyond examining the overall impact of voice AI adoption on consumption, this paper also seeks to understand the categories of products that are most affected by voice AI adoption. Previous studies have demonstrated the impact of technology adoption varies across products depending on factors such as the time and effort required for product evaluation (Park et al., 2020; Adamopoulos et al., 2020). By delving into the heterogeneous effects of voice AI adoption on the spending across different product categories, we provide deeper insights into how product features associated with the cost of information acquisition might moderate the main effect, which can assist businesses in efficiently leveraging voice AI for their marketing strategies. In addition, the impact of voice AI adoption could change over time, leaving the long-term stability of its effect on spending an unclear question. In this paper, we deepen the understanding of the effectiveness of voice AI on spending by conducting additional analyses to examine the temporal dynamic effects of the voice AI adoption. Furthermore, today's consumers have access to various digital shopping channels, including mobile, PC, and tablet, allowing them to choose the most suitable one for their specific needs and preferences. Prior literature has also examined the spillover effects of the adoption of one shopping channel on other

channels (Xu et al., 2017; Ghose et al., 2013). While voice AI adoption could enhance the overall consumption by making it more intuitive and engaging, it is also possible that consumers might shift their shopping activities from mobile and PC channels to the voice channel. Due to different comparative advantages among different channels, the spillover effects on specific channels will be contingent upon customer preferences and channel-specific characteristics. Consequently, the influence of adopting voice AI on the other two channels may depend on specific shopping contexts. The overall impact of voice AI adoption on the PC or mobile channel remains uncertain, necessitating empirical investigation for a comprehensive understanding. The findings of these heterogeneity analyses bolster the overall contribution of the paper because they further empirically verify the underlying mechanisms.

We leverage a large-scale archival dataset from Alibaba, one of the world's largest online retailers. Alibaba developed and launched Tmall Genie (or "Genie" for short), its own voice AI device similar to Amazon Alexa, offering a voice shopping experience. Our unique transactional dataset contains 31 weeks of consumer-level purchase records, including both Genie adopters and non-adopters. We examine how the introduction of Genie as a shopping channel affects consumer expenditure on the focal platform. We employ granular propensity score matching (PSM) to match adopters with non-adopters and then use the difference-in-differences (DiD) approach to identify the effect of Genie adoption on spending.

Our analyses yield several important findings. First, we find that the adoption of Tmall Genie, the voice AI device launched by the e-commerce platform, leads consumers in our sample to increase their average weekly spending on Tmall by 16.6% within the first four months after adoption. Second, we test how product characteristics moderate the effect of Genie adoption, and we provide evidence that the mechanism involves the reduction of information acquisition costs.

Third, our findings indicate a stronger Genie adoption impact on spending in product categories that necessitate less active search or comparison: specifically, product categories with low substitutability (e.g., books) and product categories which the focal consumer makes frequent purchases in and is more familiar with. This finding bolsters the overall contribution of the paper as it further empirically verifies the underlying mechanism. Fourth, we find that the overall treatment effect of Genie adoption on spending attenuates over time. For repeat purchases in the voice channel, although the positive effect weakens over time, it remains statistically significant in a relatively long term after adoption.

Moreover, we demonstrate a significantly positive spillover effect of Genie adoption on the PC channel, while observing no significant impact on the mobile channel across all the product categories. Further category-level investigation reveals the nuanced dynamics within the PC channel, wherein a positive spillover to the purchase of unfamiliar items and a negative spillover to the purchase of familiar items are observed. Within the mobile channel, voice AI leads to more purchases stemming from casual conversations with Genie, while it crowds out hands-free purchases that can be done with Genie. These findings also inform, among others, future literature on the interdependencies among shopping channels. Lastly, from a social equity perspective, we find that older consumers are equally capable of benefiting from voice AI and may be inclined towards voice-based interaction over mobile devices. Additionally, we observe that the positive spillover effects on spending are primarily due to an increase in purchase quantity rather than higher prices. Overall, we demonstrate that voice-activated shopping devices can be well-positioned to enhance the growth of the affiliated e-commerce platform.

The rest of this paper is organized as follows. In Section 2, we provide an overview of the relevant literature and highlight the contributions of this paper. Section 3 formalizes the rationale

and hypotheses behind the empirical analyses. Section 4 details the institutional background and the data. Section 5 presents the identification strategies as well as the main empirical results. We present the methods and results in testing the remaining hypotheses in Sections 6, 7, and 8 respectively. Section 9 discusses the results from post-hoc analyses and extensive robustness checks. Finally, Section 10 concludes with the implications of our study.

2. Relevant Literature

Our work contributes to several streams of relevant research, including how the adoption of digital technologies affects consumption behavior, the implications of voice AI technology, as well as how consumer behavior varies across multiple channels.

2.1 Digital Technology Adoption and Consumption

Existing literature has focused on understanding the implications of emerging digital technologies on consumer behavior and service providers. On the consumer side, for instance, Xu et al. (2023) investigated the determinants and effects of mobile payment technology adoption on consumer consumption behaviors. On the service provider side, the adoption of emerging technologies can open up new revenue streams. Pamuru et al. (2021) showed that restaurants incorporating augmented reality (AR) experiences benefit from enhanced consumer engagement and improved. Tan and Netessine (2020) studied the impact of tabletop technology on restaurant sales performance and found a great potential for introducing this technology in a large service industry that lacks digitalization. Combined, such new technologies create new avenues for consumers to engage with services, potentially altering their purchasing decisions and the outcomes for the service providers. The introduction of the mobile channel (Park et al., 2020; Ghose, 2018) and the Internet of Things (Adamopoulos et al., 2020) has been shown to be able to reshape consumer behavior and influence product sales in their respective ways. Overall, the integration of emerging

technologies has the potential to significantly impact the consumer-provider relationship, altering the way services are delivered and consumed, and how consumers interact with them. Voice AI differs from the technologies discussed in existing studies, primarily due to its unique voice modality that alters consumer behaviors in searching and purchasing (Son et al., 2023; Melumad 2023). This divergence highlights the need for further empirical investigation. Although voice AI gains attraction across various industries, the specific effects of such voice-enabled technology on e-commerce consumption remain unclear. By comprehensively examining the adoption and utilization of this novel voice AI technology, our study extends the literature on the impact of emerging technologies on consumption.

2.2 Contemporary Literature on Voice AI

Moreover, this work contributes to the expanding body of literature on the impact of voice AI, specifically in the context of e-commerce. With the escalating use of voice AI technology, understanding the impact of voice AI adoption on consumer behavior becomes increasingly vital. Despite that, we know little about how voice assistants, as a new channel for online shopping, affect e-commerce consumption. Prior literature has elaborated on the intelligent agent technology applications and adoption via theoretical framework (Kumar et al., 2016), and has demonstrated the difficulty of processing auditory information in voice commerce using experiments (Munz and Morwitz, 2022). Moreover, researchers have also conducted case studies, surveys and algorithmic designs to investigate the implications and opportunities of voice AI technology in marketing and sales policies (Chen et al., 2023; Smith, 2020; Ba et al., 2020). Additionally, Luo et al. (2019) analyzed field experiment data to study how revealing a chatbot's identity influences customer purchase decisions. Melumad (2023) studied the effects of voice technology on online information searches and found that vocalizing queries resulted in more specific and detailed descriptions, as

evidenced by survey data. Existing research on voice AI predominantly takes a descriptive, theoretical, survey or qualitative approach. Although Son et al. (2023) demonstrated the effect of smart speaker usage on consumers' digital content (namely, video on demand) searches and purchases, empirical evidence remains scarce regarding how the adoption of voice AI influences consumption on e-commerce platforms, a crucial area where voice AI is expected to play a significant role. This research aims to bridge this gap. To the best of our knowledge, this paper is among the first to empirically study the impacts of adopting voice AI on consumers' e-commerce consumption.

2.3 Channel Dynamics

The advent of the Internet and the rise of emerging technologies have broadened the array of channel options available to consumers for making purchases (Sun et al. 2021). Existing research in Information Systems and Marketing has investigated the interdependencies between various shopping channels and has demonstrated that the impact of new channel introduction can vary depending on the unique characteristics of the channels. For instance, Ghose et al. (2013) explored how consumers' online browsing behaviors vary between personal computers and mobile channels and showed that the smaller screen size of mobile phones increases the cost of browsing. Bang et al. (2013) found that the impacts of mobile channel introduction on traditional online channels depend on the fit between channel capabilities and product characteristics. In this regard, the relevant stream of literature has revealed that adopting additional shopping channels can have nuanced and multifaceted effects, depending on the specific capabilities of these channels.

Previous research has shown that adopting a new channel can lead to a complementary effect, enhancing the utility of existing channels; and a cannibalization effect, reducing the usage of existing channels. For instance, Xu et al. (2014) found a significant increase in demand at the

corresponding mobile news site in response to the introduction of a mobile channel. Xu et al. (2017) documented that tablet adoption substitutes the PC channel but complements the smartphone channel. Likewise, Xu et al. (2023) found that the mobile payment channel acted as a substitute for the offline channel and as a complement for the PC channel, and both effects increased over time. Despite the proliferation of studies on channel dynamics, no definitive tone on whether voice AI will hold a substitute or complementary relationship with the existing channels. Our work contributes to this strand of literature by evaluating how voice assistants, as a new shopping channel, affect traditional PC and mobile purchase channels.

3. Hypotheses Development

In this section, we develop four hypotheses focusing on the main effect, the moderating role of product features, dynamic effects, and channel spillover effects.

Extant literature has documented the impact of emerging technologies on consumption behavior. For example, the introduction of the mobile channel for news (Xu et al., 2014) and mobile payment (Xu et al., 2023), has substantially increased demand and consumption. The adoption of mobile shopping has been found to increase search activities and sales of leading products (Park et al., 2020). Mobile devices such as smartphones and tablets, offering convenient access, encourage consumers to incorporate mobile shopping into their habitual routines. Xu et al. (2016) found that the introduction of tablets can spur casual browsing, leading to more impulsive and diverse product purchases, thereby enhancing overall e-commerce sales. In a similar vein, the Internet of Things (IoT) has also demonstrated a significant increase in sales due to its ability to enhance convenience and reduce intangible transaction costs for consumers in purchase processes (Adamopoulos et al., 2020). As voice AI technology develops and becomes more prevalent in a range of industries, it is likely to have a similar impact on consumption and spending. Voice AI,

enabling hands-free shopping, offers greater convenience and flexibility. The enhanced user experience could lead to increased usage and higher spending on some particular product categories from consumers. Overall, we expect the adoption of voice AI will have a positive overall impact on consumer spending on the focal platform. Based on these considerations, we propose the following baseline hypothesis:

HYPOTHESIS 1 (H1). *Voice AI adoption has a positive overall effect on consumer spending.*

Despite the wide adoption of voice-enabled agents for their convenience, several attributes of the voice shopping channel can affect the types of products in which the increase in consumer spending is reflected. Research on auditory consumer behavior indicates that information presented through voice is generally more difficult to process than information presented through text or visuals (Munz and Morwitz, 2022). Given the perceived higher cognitive cost of searching and comparing unfamiliar products through voice AI, consumers may be deterred from exploring a wider range of options and instead only focus on familiar products. Consumers also have different attitudes towards convenience goods and other shopping goods, with convenience goods being those that are purchased repeatedly and require minimal search effort, while others require a more serious process of selection (Alexander, 1948). Park et al. (2020) showed that for convenience goods, which are low in search needs due to repeated purchases and enough prior knowledge, the adoption of mobile channels led consumers to choose the mobile channel as the primary means for their searches and purchases. Similarly, Adamopoulos et al. (2020) demonstrated that the introduction of IoT technologies as a sales channel is more effective for less substitutable products, which are goods that are less differentiated and require less time and effort from consumers to thoroughly evaluate their options, validating that consumers value convenience and efficiency more when shopping for these types of products. Consequently, consumers are

likely to prefer using Genie for goods that they purchase repeatedly, as it saves the time and effort needed for reorder these items using voice commands rather than having to search and comparison. Hence, we anticipate Genie to be better suited for repeat purchases, familiar product categories, and those with low substitutability, where the cost of information processing is lower for consumers. We thus propose:

HYPOTHESIS 2 (H2). The positive effect of voice AI adoption on spending through the voice shopping channel is stronger for products which do not bear high cost of information acquisition, including repeat purchase, more familiar and less substitutable product categories.

It is also worth assessing whether the positive effect of voice AI adoption on consumption remains stable over time. This could provide insight into the factors that may influence consumer spending decisions when using voice shopping. For instance, in the early stages of adoption, consumers may be hesitant in their usage of Genie due to unfamiliarity or skepticism, but as they become more familiar with it and see its benefits, they may become more inclined to use it, leading to a positive effect that increases over time. However, if consumers view voice AI as too complex or lacking, the positive effect on consumption could diminish over time. Additionally, as people become more accustomed to the technology, the novelty effect may wear off, causing the positive effect on consumption to level off.

Novelty, which can be thought of as an innate human preference for something new, unique and different from the familiar, can be created by the freshness and innovation of information technology (Masseti, 1996; Huang, 2003). Manchanda et al. (2015) described the novelty effect as a temporary positive response to the adoption of something new, but that this response tends to decline over time as people revert to their normal purchase behavior. Previous research has found

that the novelty of the technology and its ability to perform its intended services can initially be satisfying, but as the novelty wears off, consumer expectations may increase, leading to a decrease in satisfaction with the technology (Meuter et al., 2000). When consumers are disappointed with the technology's perceived usefulness and compatibility with their everyday activities, their interest and utilization of the service will decrease (Parthasarathy et al., 1998). This diminishing engagement has been evident in past studies on the effects of new technology adoption in education, where the novelty effect decreases as students get familiar with the technologies, resulting in diminished motivation to use the technology. Early studies have also demonstrated that the novelty effect of robots can quickly wear off and people have fewer interactions with them over time (Adamopoulos et al., 2020). Therefore, as the novelty of voice AI technology naturally wanes over time, it may lead to a decrease in its perceived usefulness and a corresponding decline in consumers' motivation and usage. The adoption of a new technology can be a complex process and the impact of Genie adoption may be influenced by various factors that can change over time. In this context, we reason that the ease of use and conversational nature of voice AI Genie may make the novelty effect more significant in determining the temporal dynamics of the treatment effect. Hence, we anticipate that the positive effect of Genie adoption will attenuate over time, leading to the following hypothesis:

HYPOTHESIS 3 (H3). *The positive effect of voice AI adoption diminishes over time.*

As Genie provides an additional channel for consumer interaction, adopters' shopping behavior through other channels may change. It is thus important to evaluate how the other shopping channels (namely, PC and mobile) are impacted, if any. Xu et al. (2017) found that tablet adoption substitutes desktop commerce but complements the smartphone channel. Though tablets share functional similarities with PCs, tablets offer a more portable and intuitive browsing

experience, thereby substituting both browsing and purchase activities away from the PC channel. Meanwhile, because smartphones' smaller sizes and ubiquitous connectivity augment the shortcomings of tablets, which are more location-restrictive and bulkier, users make more purchases on smartphones with fewer page browses after tablet adoption.

Likewise, as different channels have different comparative advantages, the specific channel spillover impacts will depend on customer preferences and the channel-specific traits. While mobile devices have several advantages such as portability and accessibility, which can help consumers fulfill their transactions more efficiently (Kleijnen et al., 2007), the PC channel makes it easier for consumers to search and browse products as its large screen can display more product information (Ghose et al., 2013). Given the comparative advantages in different contexts, the impact of voice AI on mobile channels could go either way. In the case where the adopters want to make a decision after a casual conversation with voice AI, but are constrained by the high cost of searching and comparing products using the voice channel, they would direct themselves to mobile devices to complete the purchase process (Xu et al., 2017). On the other hand, voice AI allows hands-free shopping (Bahmani et al., 2022), where adopters can make purchases while multitasking on mobile phones, producing a negative spillover on the mobile channel.

The spillover effect on the PC channel can also be mixed. On the one hand, when it comes to unfamiliar products consumers may switch to PC from earlier voice-initiated sessions to continue their shopping due to its expanded visual experience (Xu et al., 2017). On the other hand, for familiar items, consumers do not need much search and comparison but instead can easily use voice AI to complete the transaction, resulting in a cannibalization (negative) effect on the PC channel (Park et al., 2020). Thus, the impact of voice AI adoption on the other two channels should

depend on the specific shopping context. Consequently, the overall net effect of the Genie adoption on the PC or mobile channel remains unclear and calls for an empirical examination. Therefore:

HYPOTHESIS 4 (H4). The spillover effects of Voice AI adoption on spending in the PC and mobile channels can be either positive or negative, depending on the shopping context (e.g., the product category).

4. Institutional Background and Data Description

We acquired access to a large archival data set of consumer-level purchase data from Tmall, the online business-to-consumer retail site of the world's largest e-commerce platform, Alibaba. Since its inception in April 2008, Tmall has acquired more than 726 million registered users. In 2019, Tmall's sales revenue exceeded RMB 5.7 trillion (US \$853 billion),⁸ accounting for 55.9% of online retail sales in China.⁹ Traditional online shopping channels at Tmall include PC and mobile. In July 2017, Alibaba launched Tmall Genie, an AI-powered smart device, to link customers with new shopping experiences, smart home appliances, and other new services. As with similar voice assistant products developed by other tech giants, Genie¹⁰ enables consumers to use voice commands to make purchases. More than one million orders were placed and paid via Genie's voice-shopping feature during Alibaba's "11.11" Global Shopping Festival in 2019.¹¹ Based on the early success of Genie, Alibaba announced a plan to invest RMB 10 billion (US \$1.4 billion) in the AI system for voice assistants in May 2020.¹²

⁸ Cite from Alibaba: https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf.

⁹ Cite from press: <https://www.iimedia.cn/c1061/71838.html>.

¹⁰ We use "Genie" as a shorthand for "Tmall Genie," the focal voice-AI device.

¹¹ Cite from press: <https://www.alizila.com/millions-in-china-tapped-voice-shopping-during-11-11-tmall-genie/>.

¹² Cite from press: <https://economictimes.indiatimes.com/news/international/business/alibaba-to-invest-1-4-billion-in-ai-system-for-smart-speakers/articleshow/75842991.cms?from=mdr>.

A typical conversation between a consumer and Genie in the voice shopping context begins with the customer’s inquiry. In the illustrative example below, a consumer wants to buy a mobile refill card and does so through an efficient conversation with Genie:¹³

<i>Customer</i>	“Tmall Genie, I’d like to order a mobile refill card.”
<i>Genie</i>	“Master, I would recommend China Mobile’s refill card. The total price is 100 RMB. It will be delivered to (address). May I place the order for you?”
<i>Customer</i>	“Yes, please place the order.”
<i>Genie</i>	“Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the ‘beep,’ say ‘Tmall Genie, 2065.’” (Here, 2065 is the authentication code randomly generated by the system.)
<i>Customer</i>	“2065.”
<i>Genie</i>	“Alipay discount is applied. If you want to know the delivery status, you can let me know by saying ‘Tmall Genie, tracking information.’”

We collaborate with Alibaba to collect data from 38,179 randomly selected consumers. Our data consists of 5.59 million consumer-level transactions from April 1 to October 31, 2019; 5.23 million purchases were made on mobile devices (including smartphones and tablets), while 0.35 million were made on PCs, and 0.005 million were made with voice shopping channel.¹⁴ The purchase transaction data includes an anonymized user ID, product category name (Tmall offers 137 product categories; see Appendix A4), purchase date, confirmed payment amount, the access channel (PC, mobile, or voice), the consumer’s age and gender, and the consumer’s number of historical purchases prior to the first day of data collection (April 1, 2019).

4.1. Sample of Genie Adopters and Non-Adopters

In 2019, Alibaba reported that over 10 million Genie units had been activated, rendering Genie the most popular voice assistant product in China.¹⁵ With the above-described data, we leverage a small set of Genie adopters to conduct the analysis. In our sample, “adopters” are customers who

¹³ We provide two additional examples of shopping-related conversations with Genie in Appendix A2.

¹⁴ Given the small fraction of voice-assisted purchases in our dataset, our estimated effects may be most applicable to the phenomenon of early penetration.

¹⁵ Cite from SEC: https://www.sec.gov/Archives/edgar/data/1577552/000110465920082881/a20-6321_46k.pdf.

adopted Genie during the first week of July (July 1–7), 2019, while “non-adopters” are customers who never adopted Genie at any time before the end of the sample period (October 31, 2019). Our dataset has 693 adopters that constitute all the users whose first adoption of Genie took place in the first week of July (i.e., from July 1 to July 7, 2019), a similar procedure adopted by Xu et al. (2023).¹⁶ The remaining 37,486 control users were randomly drawn from all the users (based on their user IDs) who had not adopted Genie at any time before the end of the sample period (October 31, 2019). We place no other requirements (e.g., the number of purchases exceeds a specific threshold) for consumers to be selected into our sample, ensuring no sample selection bias in our data construction that may undermine the external validity of our findings. In the post-treatment period (i.e., after the first week of July), adopters could make purchases on Tmall via all three channels (PC, mobile, and voice), while non-adopters were restricted to the PC and mobile channels only.

One may wonder whether some consumers chose to adopt Genie due to a non-random shock such as a promotion event. To address this concern, we confirmed with Alibaba that no promotion activities occurred during the first week of July 2019.¹⁷ However, the identification of the impact of a consumer’s Genie adoption on the consumer’s Tmall spending remains challenging because self-selection creates endogeneity concerns. For instance, if adopters have more disposable income than non-adopters, then adopters might be expected to make more purchases on Tmall; the confounding of disposable income would lead to an overestimation of the treatment effect. We utilize various econometric techniques to address the potential endogeneity concerns, detailed in Section 5.

¹⁶ Our data does not include those whose first-time adoption happened in other weeks. In our data, all adopters adopted Genie only once. We provide model-free evidence on the adoption rate by gender and by age group in Appendix A3.

¹⁷ We further discuss how we address the temporal selection concern in subsection 9.3.2.

4.2. Measuring Spending

We measure consumer-level spending as the *total amount spent on Tmall* (the confirmed payment amount across all 137 product categories) and the *amount spent in each product category*. We exclude the purchase of Genie itself from the calculations; otherwise, the outcome variables for adopters and non-adopters would not be comparable in the first place. We aggregate the raw data at the user-week level, reasoning that spending would fluctuate too much at the user-day level.

Table 1. Summary Statistics¹⁸

Variable	N	Mean	SD	Min.	Max.
User Demographics					
<i>Age</i>	38,179	31.89	7.656	18	57
<i>Gender</i> (male = 1)	38,179	0.583	0.493	0	1
<i>Number of Historical Purchases</i>	38,179	301.2	461.8	0	13,641
Spending per Category					
<i>Spending_Mobile Refill Card</i>	1,183,549	11.51	39.52	0	2,994
<i>Spending_Transportation</i>	1,183,549	2.803	54.89	0	9,368
<i>Spending_All others</i>	1,183,549	1.853	55.46	0	24,000
<i>Spending_Movies & Sports Ticket</i>	1,183,549	1.220	11.92	0	1,656
<i>Spending_Fresh Hema</i>	1,183,549	1.152	15.47	0	2,522
<i>Spending_O2O Store</i>	1,183,549	1.032	31.83	0	6,068
<i>Spending_Hotel</i>	1,183,549	0.414	18.30	0	6,042
<i>Spending_Game Time Cards</i>	1,183,549	0.285	20.41	0	7,001
<i>Spending_Used Good</i>	1,183,549	0.185	14.32	0	4,300
<i>Spending_Tencent QQ</i>	1,183,549	0.147	13.10	0	8,645
<i>Spending_Mobile Phone Plans</i>	1,183,549	0.087	3.734	0	1,980
<i>Spending_Video Game</i>	1,183,549	0.066	9.053	0	5,005
<i>Spending_Insurance</i>	1,183,549	0.035	3.084	0	1,435
<i>Spending_Coffee & Beverages</i>	1,183,549	0.015	0.998	0	294
<i>Browsing_SmartDevices</i>	1,183,549	0.080	2.474	0	832

Note: Summary statistics are calculated on the unmatched sample of 693 adopters and 37,486 non-adopters, observed over 31 weeks from April 1 through October 31, 2019. The unit of analysis is the individual user for demographics and the user-week for the purchase variables.

¹⁸ Throughout the paper, we use 4 significant figures (e.g., 1.000, 10.00, and 100.0) to report the results.

Our data collection period (April 1, 2019 – October 31, 2019) corresponds to 31 weeks (i.e., from week 14 through week 44 of 2019), so the panel dataset contains 1,183,549 observations from 38,179 consumers. Of these 31 weeks, the first 13 weeks (April 1 – June 30, 2019) constitute the pre-treatment period, 1 week (July 1 – July 7, 2019) constitutes the treatment period, and 17 weeks (July 8 – October 31, 2019) constitute the post-treatment period.

We report the summary statistics in Table 1. For clarity in presentation, we report statistics solely for the top 13 product categories, aggregating spending in the remaining 124 categories under the “All others” category.

5. Main Effect

5.1. Identification Strategy

Our goal is to estimate the average treatment effect on the treated (ATT), namely, the effect of a consumer’s Genie adoption on their spending on Tmall. A major challenge we face is the lack of a randomized assignment of consumers into treatment and control groups. We use two main strategies to address this challenge. First, in the main text, we estimate a difference-in-differences (DiD) model, a widely adopted approach that eliminates persistent linear and additive individual-specific effects that may introduce endogeneity. To construct an appropriate sample for the DiD estimation, we attempt to control for the unobserved need for spending in the absence of Genie by conditioning the sample on a rich set of observed characteristics (for a similar approach, see Datta et al., 2018; Bronnenberg et al., 2010). Specifically, we use a quasi-experimental matching procedure (detailed in Section 5.3) in which we match adopters with similar non-adopters based on each user’s propensity to adopt Genie, as estimated from the user’s demographics and granular shopping behaviors. Moreover, consumers inclined to purchase Tmall Genie may also tend to spend more time examining smart device product pages, potentially leading to a reverse causality

issue. To rule out this potential endogeneity concern, we introduce an additional matching covariate – the browsing history of smart devices – adding an additional dimension to pair customers from both treatment and control groups. Holding all other observable characteristics equal, individuals with a similar number of pageviews on the smart device product page are more likely to have the same level of interest in purchasing it.¹⁹ Additionally, we incorporate individual and time-fixed effects in the panel regressions to control for unobservable factors, and undertake a series of checks to ensure the robustness of our findings. Moreover, in Appendix A13, we strengthen the main finding with an instrumental variable approach.

5.2. Comparison of Adopters and Non-adopters

First, we compare the demographics and purchasing behaviors of adopters and non-adopters during the pre-treatment period. Table 2(a) shows that, on average, adopters and non-adopters differ significantly: Adopters are younger than non-adopters (31.0 years old vs. 31.9 years old, $p = 0.002$), and the adopter group has a smaller share of males (39.0% vs. 58.7%, $p < 0.001$). Furthermore, in the pre-treatment period, adopters spent more per week than non-adopters (RMB 34.35 vs. RMB 19.36, $p < 0.001$) and had more historical purchases (560.0 vs. 296.4, $p < 0.001$). Regarding the browsing history of smart devices, adopters have a higher number of pageviews per week compared to non-adopters during the pre-treatment period (0.324 vs. 0.076, $p < 0.001$). These significant pre-treatment differences between the treatment and control groups might explain some or all of the differences in spending in the post-treatment period, thus posing a challenge to the clean identification of Genie adoption's effects. In the next section, we use quasi-experimental methods to address self-selection.

¹⁹ We thank the anonymous review team for raising this concern and providing a way to address it.

5.3. Propensity Score Estimation and Matching

To mitigate significant differences between the treatment and control groups and thereby alleviate the self-selection issue, we match each adopter with non-adopters who have a similar propensity of being treated, based on similarities in observable characteristics (while including individual fixed effects could partially address the selection issue, this approach imposes a linear functional form, whereas matching allows for more flexibility). Our matching utilizes one-to-five matching with replacement to derive the closest matched nonadopters. As a result, each adopter is paired with non-adopters who have similar propensities of being treated. This approach enables a fair comparison of shopping behaviors between these two groups of consumers. In particular, we estimate each individual's adoption propensity as a function of observed variables:

$$Pr(Adoption_i) = Pr(\varphi_0 + \varphi^T Z_i + \epsilon_i > 0) \quad (1)$$

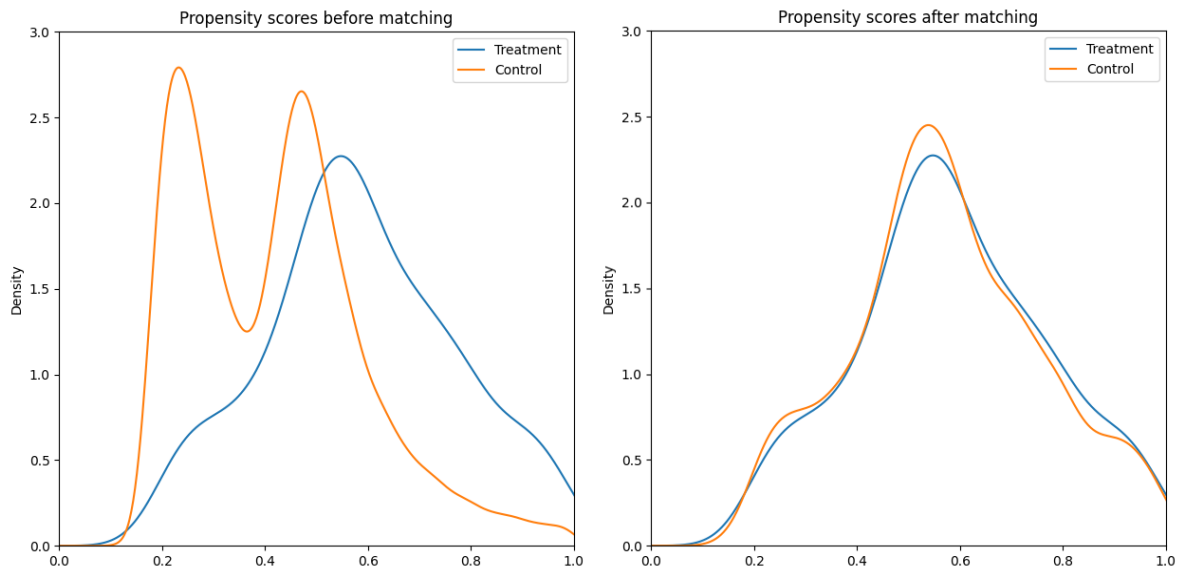
where Z_i is a vector of observed individual-specific characteristics, φ is a vector of coefficients, φ_0 represents the constant term, and ϵ_i represents the error term. We assume that ϵ_i are independent and identically distributed random variables with a type I extreme value distribution, making the propensity score estimation a logit model.

The covariates in Z_i describe the user's demographics (age and gender) and pre-treatment purchase behavior (the number of historical purchases, spending in each product category, and browsing history of the smart devices). Note that we included highly granular purchase behavior data (at the product category level) to ensure that the matched sample contains adopters and non-adopters whose purchase behaviors are as similar as possible. However, not all of the 137 product categories defined by Alibaba had sufficient non-zero observations for matching,²⁰ so we used the

²⁰ Specifically, we needed at least 285 purchase records per category for matching.

13 categories with the highest purchase frequency and aggregated the remaining 124 categories (labeled "all others"), as shown in Table 2. Thus, we used consumer purchase behaviors in 14 categories to conduct matching.²¹

Figure 1. Distribution of Propensity Scores Before and After Matching



Note: Logit model with standard errors in parentheses. Estimates are calculated on an unmatched sample of 693 adopters and 37,486 non-adopters in the pre-treatment period; the unit of analysis is the individual user. The dependent variable is whether the user adopted Tmall Genie (adoption = 1) during the first week of July, 2019, or did not adopt Genie at any time before the end of the sample period (adoption = 0).

²¹ As a robustness check, we redo the matching procedure with four less granular categories, and we re-estimate the DiD model on the new matched sample. The results are qualitatively the same. See details in Section 9.4.3.

Table 2. Comparison of Adopters and Non-Adopters Before and After Matching

	Adopter (N)	Adopter (Mean)	Non-adopter (N)	Non-adopter (Mean)	p-value of t-test
<i>(a) Before matching</i>					
<i>Age</i>	693	31.00	37,486	31.90	0.002
<i>Gender (Male = 1)</i>	693	0.390	37,486	0.587	0.000
<i># Historical Purchases</i>	693	560.0	37,486	296.4	0.000
<i>Spending_Mobile Refill Card</i>	693	17.00	37,486	11.52	0.000
<i>Spending_All others</i>	693	3.722	37,486	1.963	0.004
<i>Spending_Transportation</i>	693	2.520	37,486	1.448	0.033
<i>Spending_Movies & Sports Ticket</i>	693	2.348	37,486	0.992	0.000
<i>Spending_Fresh Hema</i>	693	2.123	37,486	1.032	0.003
<i>Spending_Game Time Cards</i>	693	2.118	37,486	0.305	0.000
<i>Spending_O2O Store</i>	693	1.665	37,486	1.101	0.269
<i>Spending_Hotel</i>	693	1.052	37,486	0.389	0.009
<i>Spending_Used Good</i>	693	0.606	37,486	0.187	0.040
<i>Spending_Tencent QQ</i>	693	0.467	37,486	0.197	0.387
<i>Spending_Mobile Phone Plans</i>	693	0.167	37,486	0.099	0.229
<i>Spending_Video Game</i>	693	0.137	37,486	0.067	0.632
<i>Spending_Insurance</i>	693	0.096	37,486	0.040	0.101
<i>Spending_Coffee & Beverages</i>	693	0.009	37,486	0.015	0.629
<i>Browsing_SmartDevices</i>	693	0.324	37,486	0.076	0.000
<i>(b) After matching</i>					
<i>Age</i>	693	31.00	3,192	31.07	0.816
<i>Gender (Male = 1)</i>	693	0.390	3,192	0.408	0.382
<i># Historical Purchases</i>	693	560.0	3,192	556.2	0.899
<i>Spending_Mobile Refill Card</i>	693	17.00	3,192	16.88	0.898
<i>Spending_All others</i>	693	3.722	3,192	4.503	0.579
<i>Spending_Transportation</i>	693	2.520	3,192	2.992	0.646
<i>Spending_Movies & Sports Ticket</i>	693	2.348	3,192	2.342	0.988
<i>Spending_Fresh Hema</i>	693	2.123	3,192	1.858	0.630
<i>Spending_Game Time Cards</i>	693	2.118	3,192	0.942	0.214
<i>Spending_O2O Store</i>	693	1.665	3,192	1.974	0.663
<i>Spending_Hotel</i>	693	1.052	3,192	1.298	0.717
<i>Spending_Used Good</i>	693	0.606	3,192	0.318	0.244
<i>Spending_Tencent QQ</i>	693	0.467	3,192	0.165	0.049
<i>Spending_Mobile Phone Plans</i>	693	0.167	3,192	0.249	0.368
<i>Spending_Video Game</i>	693	0.137	3,192	0.164	0.869
<i>Spending_Insurance</i>	693	0.096	3,192	0.076	0.629
<i>Spending_Coffee & Beverages</i>	693	0.009	3,192	0.009	1.000
<i>Browsing_SmartDevices</i>	693	0.324	3,192	0.224	0.224

Note: Calculated over the pre-adoption period.

5.4. Main Effect using Difference-in-Differences

We estimate the difference-in-differences (DiD) model on the matched sample to quantify the effect of a consumer's Genie adoption on the consumer's spending on the focal e-commerce platform. We estimate the main effect using the following equation:

$$y_{it} = \gamma \cdot (Adoption_i \times Post_t) + \delta_i + \tau_t + \varepsilon_{it} \quad (2)$$

where y_{it} is the total amount spent by consumer i in week t , γ captures the treatment effect, δ_i is consumer-level fixed effects, τ_t is week-level fixed effects, and ε_{it} is the random error term. We report robust standard errors clustered at the user level. In summary, we combine the propensity score matching method (to select non-adopters who are like adopters) and the DiD approach. We assume that the unobserved ε_i in the propensity score model is independent of the unobserved ε_{it} in the DiD regression model.

Table 3. Main Effect

	<i>All</i> <i>Spending (RMB)</i>	<i>Mobile</i> <i>Spending (RMB)</i>	<i>PC</i> <i>Spending (RMB)</i>
Adoption \times Post	5.719** (2.018)	0.795 (2.510)	2.157*** (0.690)
Constant	32.92*** (0.435)	27.943*** (0.260)	4.902*** (0.071)
Individual FEs	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes
Observations	120,435	120,435	120,435
Adjusted R ²	0.138	0.103	0.058

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table 3 reports the estimated main effect of a consumer's Genie adoption on their weekly spending amount. The results indicate that Genie adoption leads the average adopter to spend RMB 5.719 more per week. Correspondingly, within the first four months after adoption, Genie led to an average increase of 16.6% in weekly spending for adopters in our sample, compared to the pre-

treatment mean of RMB 34.35. The finding of a significant positive effect of Genie adoption on Tmall spending implies that having Genie increases the consumer's tendency to place orders. These results confirm our Hypothesis 1 that voice AI adoption has a positive overall effect on consumer spending.

To delve deeper into what drives the increase in spending, we also estimate the impact of Genie adoption on spending in the PC and mobile channels. As shown in columns (2) and (3) in Table 3, Genie adoption has an insignificant impact on spending in the mobile channel and a significantly positive impact on spending in the PC channel. The residual can thus be attributed to the voice channel. Section 8 delves deeper into the effects of channel dynamics.

6. Moderating Effect of Product Category Features

This section delves into the moderating role of product category features on the effect of Genie adoption on spending through the voice channel. These analyses have both practical significance (for the strategic placement of Genie promotions) and theoretical significance (by offering insight into the underlying mechanism). Specifically, we investigate three potential moderators at the product-category level: new versus repeat purchases, product substitutability, and familiarity.

The choices of moderators are informed by the finding that information presented by voice can be more difficult to process than the same information presented in writing (Munz and Morwitz, 2022), leading to a higher cost of information acquisition in the voice shopping channel. Products that are familiar, less substitutable and purchased repeatedly do not require extensive searching or comparison, which makes the shopping process less costly (Alexander, 1948; Park et al., 2020; Adamopoulos et al., 2020). We thus anticipate Genie to be better suited for purchasing these types of products.

To define new versus repeat purchases, we set a pre-study window of 6 weeks and identify the set of categories from which the consumer has made purchases (denoted as Set A).²² For the pre-adoption period, we define repeat purchases as those in Set A and new purchases as those out of Set A (denoted as Set B). For the post-adoption period, we then define repeat purchases as those in either Set A or Set B and new purchases as those out of both Set A and Set B (denoted as Set C).²³ We expect the cost associated with ordering repeat purchases will be lower than that for new products, as it costs consumers less time in searching or comparing. We re-estimate Eq. (2), with the spending on new and repeat purchases as two separate dependent variables.

Table 4. Moderating Role of New vs Repeat Purchases on Main Effect Across All Channels

	<i>All Channels (RMB)</i>		<i>Voice Channel (RMB)</i> ²⁴		
	<i>New Purchase</i>	<i>Repeat Purchase</i>		<i>New Purchase</i>	<i>Repeat Purchase</i>
Adoption × Post	0.119 (1.364)	5.600*** (1.492)	Post	0.412*** (0.048)	2.783*** (0.167)
Constant	12.24*** (0.294)	20.68*** (0.321)	Constant	0.000 (0.028)	0.000 (0.070)
Individual FEs	Yes	Yes	Individual FEs	Yes	Yes
Week FEs	Yes	Yes	Week FEs	No	No
Observations	120,435	120,435	Observations	21,483	21,483
Adjusted R ²	0.036	0.118	Adjusted R ²	0.022	0.031

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

²² We also set the pre-study window as 3, 4, 5, 7, and 8 weeks and obtain consistent results. The corresponding results are reported in the robustness check section 9.4.5 and Appendix A11.

²³ An alternative way is to define the new purchase as the first item purchased in each product category and the repeat purchase as the subsequent purchases in each category. For instance, if user i in week t purchases a product of category j , which has never been purchased before, the amount spent on this product counts towards the spending on new purchase, whereas when the same user i later purchases any product of category j , the amount spent counts as spending on repeat purchase. However, this could result in selection bias that over time, there are fewer (more) categories from which the purchases would qualify for new (repeat) purchases. We thank an anonymous reviewer for pointing out this problem and suggesting a method to address it.

²⁴ The spending in the voice channel for the control group cannot be observed. Therefore, we only keep the observations for the treatment group and the estimates capture the average spending in the voice channel for the treated users. For other tables involving the spending in the voice channel, a similar model specification ($y_{it} = \gamma \cdot Post_t + \delta_i + \varepsilon_{it}$) is applied. We thank an anonymous editor for making this point.

In Table 4, we find a significantly positive effect on repeat purchases while no significant effect on new purchases for both the overall spending and the voice channel spending, indicating that the positive treatment effect is primarily driven by consumers' repeat purchases, supporting our Hypothesis 2 that the effect of voice AI adoption on spending through the voice channel is stronger for products that do not bear the high cost of information acquisition.²⁵

To further test Hypothesis 2, we leverage two additional moderators, product substitutability and familiarity, using data at the user-week-category level.

Substitutability describes the extent to which consumers perceive products as interchangeable with each other. If a product category has *low substitutability*, then each product in the category can be treated as an independent market. Products with low substitutability should not require much active searching or comparison, and can therefore be more compatible with voice shopping. We anticipate that the positive effect of Genie adoption on spending is stronger for categories with low substitutability.

We determine which product categories with *low substitutability* by reviewing existing findings. Prior literature has found that books (Chevalier and Goolsbee, 2003; McMillan, 2007), music (Shiller and Waldfogel, 2009), antiques and artworks (McAfee and McMillan, 1987) have sufficiently low substitutability that each product can be treated as an independent market. Additionally, the prescription drug category is included, given that consumers are limited to

²⁵ We acknowledge our data limitation that we don't have access to historical data before the data collection period. The way we define new and repeat purchases is thus subject to this limitation. Given that the entire purchase history is not available, the number of categories in Set B and Set C could be larger than that using the entire purchase history. Such a data limitation has two implications for new and repeat purchases: 1) The effect of voice AI adoption on new purchases could be overestimated. Nevertheless, we find insignificant treatment effects on new purchases across different pre-study time window specifications. Therefore, the actual effect on new purchases should still be insignificant. 2) The effect of voice AI adoption on repeat purchases could be underestimated. That is, the actual effect on repeat purchases should be larger. If the entire purchase history can be provided, the difference between the effects on new purchases and repeat purchases could be greater, further supporting our Hypothesis 2 that the effect of voice AI adoption on spending through the voice channel is stronger for products that do not bear high cost of information acquisition. We thank an anonymous reviewer for suggesting this discussion.

choosing the product prescribed to them. We estimate the moderating effect of low substitutability with the following equation:

$$y_{ijt} = \gamma \cdot Post_t + \lambda_{LS} \cdot (Post_t \times LowSubstitutability_j) + \delta_i + \sigma_j + \varepsilon_{ijt} \quad (3)$$

where y_{ijt} is the amount spent by consumer i in week t on category j ; σ_j is the category fixed effects; $LowSubstitutability_j$ is a dummy variable that equals 1 if j belongs to a category with low substitutability and 0 otherwise; γ captures the main effect; and λ_{LS} represents the interaction effect of Genie adoption and low substitutability.

Familiarity, like *low substitutability*, should enable consumers to make purchases without an active search or comparison because consumers would already possess prior knowledge about the product options in the category. We proxy *familiarity* with the number of times a product category j was purchased by consumer i in the pre-treatment period.²⁶ We anticipate that the familiarity of purchases in a product category positively moderates the effect of Genie adoption on spending in that category, and estimate the moderating effect with the following equation:

$$y_{ijt} = \gamma \cdot Post_t + \lambda_F \cdot (Post_t \times Familiarity_{ij}) + \delta_i + \sigma_j + \varepsilon_{ijt} \quad (4)$$

where y_{ijt} is the amount spent by consumer i in week t on category j ; $Familiarity_{ij}$ is the number of times a product category j was purchased by consumer i in the pre-treatment period as defined above; γ captures the main effect; and λ_F represents the interaction effect of Genie adoption and familiarity.

To test how substitutability and familiarity of a product category moderate the effect of Genie adoption on spending in the focal category, we estimate Eq. (3) and Eq. (4) on the purchase data in the voice channel and present the findings in the first two columns of Table 5. Moreover,

²⁶ We thank an anonymous reviewer for suggesting this measurement.

we also include both moderators in the same model and report the result in the third column of Table 5.²⁷

The estimated coefficients of the interaction terms $Post \times Low\ Substitutability$ and $Post \times Familiarity$ are both positive and significant, supporting our expectation that the impact of Genie adoption on spending through the voice channel is stronger in product categories that are less substitutable and more familiar. These results corroborate with the findings in Table 4 to jointly confirm Hypothesis 2.

Table 5. Heterogenous Effect across Product Category Features

		Voice Channel Spending (RMB)	
Post	0.000 (0.001)	0.015*** (0.001)	-0.002 (0.001)
Post \times Low Substitutability	0.823*** (0.008)		0.654*** (0.008)
Post \times Familiarity		0.259*** (0.001)	0.202*** (0.001)
Constant	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Individual FEs	Yes	Yes	Yes
Category FEs	Yes	Yes	Yes
Observations	2,943,171	2,943,171	2,943,171
Adjusted R ²	0.026	0.026	0.028

Note: Robust standard errors are in parentheses.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

7. Novelty Effect

Does the positive effect of Genie adoption on spending remain stable over time? To investigate the temporal dynamics of Genie adoption, we estimate the treatment effect in three post-treatment

²⁷ We thank an anonymous reviewer for this suggestion. Besides, the correlation coefficient between substitutability and familiarity is 0.3395.

periods: short term (post-treatment week 1, λ_{ST}), medium term (post-treatment weeks 2–7, λ_{MT}), and long term (post-treatment weeks 8–17, λ_{LT}). Following Zhou et al. (2020) and Datta (2018), the three effects on the overall spending and spending in the voice channel are captured using the following equation:

$$y_{it} = \lambda_{ST} \cdot (Adoption_i \times Post_t \times I[0 \leq week_since_adoption_t \leq 1]) + \lambda_{MT} \cdot (Adoption_i \times Post_t \times I[2 \leq week_since_adoption_t \leq 7]) + \lambda_{LT} \cdot (Adoption_i \times Post_t \times I[week_since_adoption_t \geq 8]) + \delta_i + \tau_t + \varepsilon_{it} \quad (5)$$

where y_{it} is the total amount spent by consumer i in week t , and $I[week_since_adoption_t]$ is the indicator variable. In Table 6, we report the estimated treatment effects in the short, medium, and long term on the overall spending in Column (1). We also re-estimate Eq. (5) to observe the changes in treatment effect over time while taking into consideration the moderating effect of product features, with the subsets of new (Column 2) and repeat purchases (Column 3) in the voice shopping channel.

Table 6. Novelty Effect

	(1) All Total Spending (RMB)		(2) Voice ²⁸ New Purchase (RMB)	(3) Voice Repeat Purchase (RMB)
Adoption × Post × Short-term	10.17* (4.211)	Post × Short-term	2.263*** (0.355)	8.676*** (0.599)
Adoption × Post × Medium-term	7.796** (2.736)	Post × Medium-term	0.252*** (0.068)	2.893*** (0.219)
Adoption × Post × Long-term	3.583 (2.332)	Post × Long-term	0.137*** (0.037)	1.538*** (0.207)
Constant	32.92*** (0.435)	Constant	0.000 (0.028)	0.000 (0.097)
Individual FEs	Yes	Individual FEs	Yes	Yes
Week FEs	Yes	Week FEs	No	No
Observations	120,435	Observations	21,483	21,483
Adjusted R ²	0.138	Adjusted R ²	0.013	0.052

Note: Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001

²⁸ The model specification for columns (2) and (3) is the same as Equation (5) except that Adoption = 1 and Week FEs are not included.

We find that the overall treatment effect is the strongest in the short term (+ RMB 10.17 weekly per user). The effect size diminishes in the medium term (+ RMB 7.796 weekly per user) and becomes insignificant in the long term. Breaking down the overall spending, we learn from Column (3) that the treatment effects are positive and significant throughout the post-treatment period on repeat purchase in the voice channel. The treatment effect is the strongest in the short term (+ RMB 8.676 weekly per user in the short term)²⁹, diminishes in the medium term (+ RMB 2.893 weekly per user) and diminishes further in the long term (+ RMB 1.538 weekly per user). It is likely that the novelty effect is at work - consumers who adopt Genie are initially enticed by a curiosity about the novel AI gadget and/or by an urge to get familiar with the product. As the product becomes more familiar, consumers settle into a more modest pattern of increased spending and may eventually revert to their purchase habits before Genie adoption. Moreover, there is a positive and significant impact on new purchase (Column 2) in the short term, though the effects are comparatively smaller than those on repeat purchase. The novelty of voice shopping technology to the consumers can encourage them to explore a wider range of new products but such an effect is small and decays over time, as shown by the even smaller coefficients in the medium and long terms.

Overall, our findings reveal that the voice assistant device promotes increased repeat purchases in the short run but fails to retain the same level of interest over the long run, aligning with the concept of a novelty effect. By comparing the magnitude of the effects on the total spending and the spending on the voice channel, we find that in the short term, the sum of the effects on new purchases and repeat purchases in the Voice channel is close to the effect on the

²⁹ We discovered that the majority of products purchased by consumers using Genie are of fixed demand, such as mobile refill cards. The sharp increase in short-term spending in the voice channel, which subsequently diminishes, may be due to consumers switching from alternative platforms to the focal platform. Unfortunately, without data from multiple online shopping platforms, we are unable to verify this possibility.

total spending, indicating that the boost in overall spending in the short term is mainly attributed to the increased spending in the voice channel. However, in the medium and long terms, the sum of the effects on new purchases and repeat purchases in the Voice channel cannot fully account for the effect on the total spending, suggesting that there could be a cross-channel spillover effect due to Genie adoption, which we further investigate in the next section.

8. Channel Dynamics

The findings presented so far indicate a positive effect of Genie adoption on Tmall spending. By providing a new channel for shopping, Genie has the potential to cause spillover effects on spending in the PC and mobile channels. We proceed to examine if the adoption leads to an increase in spending across channels (i.e., a complementary effect), or if it merely prompts consumers to switch their spending from the PC and mobile channels to the voice channel (i.e., a cannibalization effect). To test Hypothesis 4, we start by re-estimating Eq. (2), the main effect model, using the purchase data from the mobile and PC channels, as shown in Table 7. Additionally, we further divide the purchases into new and repeat purchases to get a more comprehensive view of the estimated treatment effect for each channel.

From Table 7, we can derive the following conclusions regarding the overall spillover effects onto the other two channels. First, there is no significant negative treatment effect, so the voice channel does not seem to be cannibalizing the other two channels. Instead, in the PC channel, we observe positive and significant effects on spending on repeat purchases (Column 4) in the medium (+ RMB 2.155, $p < 0.01$) and long term (+ RMB 1.312, $p < 0.05$), while the effect on new spending in the PC channel is insignificant (Column 3). Genie adoption has a significant positive spillover effect on repeat purchases in the PC channel starting from the medium term. We reason that as a consumer develops more trust and familiarity with the voice AI over time, one may engage

in more conversations related to personal life and interests.³⁰ In the case where the adopters want to proceed with the purchase process after the conversation with voice AI, but are constrained by the high cost of searching and comparing products using voice, they would direct themselves to other shopping channels to complete the purchase process. Particularly for the PC channel, due to its larger screen size and more thorough visual representation of product information, the positive spillover effect is mainly oriented here. Last but not least, the overall spillover effect of Genie adoption on spending in the mobile channel is not significant, regardless of new or repeat purchases. The insignificant effect could be attributed to a recognized limitation of the mobile channel: the small screen size displays fewer product options than a PC, which makes the search harder (Ghose et al. 2013).

Table 7. Heterogeneity in the Spillover Effect across Channels

	Mobile		PC	
	(1) <i>New Purchase</i> (RMB)	(2) <i>Repeat Purchase</i> (RMB)	(3) <i>New Purchase</i> (RMB)	(4) <i>Repeat Purchase</i> (RMB)
Adoption × Post × Short-term	-0.906 (2.586)	-1.108 (2.859)	0.819 (1.148)	0.850 (1.124)
Adoption × Post × Medium-term	0.604 (1.680)	1.645 (1.857)	0.679 (0.746)	2.155** (0.730)
Adoption × Post × Long-term	-1.336 (1.432)	1.821 (1.583)	0.537 (0.636)	1.312* (0.622)
Constant	9.810*** (0.267)	18.13*** (0.295)	2.381*** (0.119)	2.521*** (0.116)
Individual FEs	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435
Adjusted R ²	0.032	0.114	0.010	0.062

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

³⁰ Since all purchases start from consumer-initiated conversations and Tmall Genie cannot initiate any conversation on its own, we expect most conversation topics are those that the consumers have prior knowledge or interest in, rather than new topics which can potentially stimulate spending on new purchase, hence the insignificant spillover effect on new spending in the PC channel.

To attain a better comprehension of the underlying mechanisms engendering spillover effects, we assess treatment effects at the category level within each channel (see results in Appendix A12). Although statistically significant effects are absent in most categories, Genie adoption manifests significant effects in four specific categories, consistent with the argument posited in Hypothesis 4. Notably, within the PC channel, the adoption of Genie leads to a significant reduction in expenditure on *Mobile Refill Cards* and a significant increase in expenditure on *Movie & Sports Tickets*. This attests to the cannibalization effect of Genie adoption on the acquisition of familiar items, which reduces transaction costs for consumers (Park et al., 2020). Conversely, concerning unfamiliar items, Genie adoption exerts a complementary effect on the PC channel, where consumers benefit from an enhanced transactional experience facilitated by the expanded screen (Ghose et al., 2013).³¹ Within the mobile channel, a negative spillover effect is found in the spending on *Online Game Point Cards*, a transaction conveniently executed hands-freely through the voice channel (Bahmani et al., 2022). By contrast, there is a positive spillover effect on the spending on *Snacks, Nuts & Specialties*. This divergence emerges as consumers, after initiating dialogues with voice AI, transition to the mobile channel to complete the purchases, motivated by the high costs of product comparisons via voice (Xu et al., 2017).³² Taken together, these findings provide empirical evidence for the channel dynamics proposed in Hypothesis 4.

9. Post-hoc Analyses

To derive further insights into the potential impacts of Genie adoption, we conduct a series of post-hoc analyses to examine: 1) the heterogeneity across user demographics, 2) channel dynamics for

³¹ In Appendix A12, we provide an example of the “*Movie & Sports Tickets*” category to highlight the potential challenges in relying solely on the voice channel for transactions.

³² More detailed data on consumers' purchase journeys, particularly the search aspect, would have allowed for a deeper exploration of the complementarity mechanisms between channels. Regrettably, we are unable to obtain such finer-grained data.

middle-aged consumers, 3) the treatment effect on price and purchase quantity which may have driven the effect on spending. Additionally, we conduct several robustness checks.

9.1 Heterogeneity in the Treatment Effect across User Demographics

The impact of Genie adoption on spending potentially varies depending on consumer characteristics, notably age and gender. It is also interesting and socially meaningful to investigate and provide more theoretical substantiation for these variations, as it could provide a deeper understanding of how different segments of consumers interact with this technology and how it affects their spending patterns. We investigate the heterogeneity in the treatment effect across consumer age and gender in Appendix A5.

9.2 Middle-aged Consumers' Channel Dynamics

As presented in Appendix A5, the effect of Genie adoption on purchase spending is negative for consumers older than 45 years of age.³³ As technology continues to advance, it can be challenging for older individuals to keep up and take advantage of the latest developments. From a social equity perspective, it is important to investigate the changes behind the overall decrease in spending for this group of adopters. This can complement the factors that researchers and developers should take into consideration when designing inclusive AI systems, such as the need for accessibility, ease of use, and the ability to interact with technology in a natural and intuitive way. We select the consumers of age ≥ 45 in the treatment group together with their matched consumers in propensity score matching. We present the findings in Appendix A6.

³³ The maximum age in our sample is 57. According to the definition from Wikipedia (https://en.wikipedia.org/wiki/Middle_age), those older than 45 but below 60 belong to the "middle age" group.

9.3 Total Quantity and Average Price

To more thoroughly understand the mechanism driving increased spending, we expand our main analysis to include two additional dependent variables: the total quantity and the average price of items purchased by users. This approach will enable us to determine whether the increased spending is driven by consumers buying more items, at a higher price point, or a combination of both factors.

We first examine the impact of Genie adoption on the average price of products purchased across shopping channels over time. In Table 8, the estimated treatment effects on the average price of products purchased in mobile and PC channels are insignificant in general (Columns 1 to 4). This indicates that Genie adoption does not lead consumers to purchase products of higher prices.

Table 8. Treatment Effect on Average Price

	Average Price (RMB)					
	Mobile		PC		All	
	(1) <i>New Purchase</i> (RMB)	(2) <i>Repeat Purchase</i> (RMB)	(3) <i>New Purchase</i> (RMB)	(4) <i>Repeat Purchase</i> (RMB)	(5) <i>New Purchase</i> (RMB)	(6) <i>Repeat Purchase</i> (RMB)
Adoption × Post × Short-term	-0.070 (1.473)	-1.267 (1.611)	0.830 (0.766)	0.150 (0.861)	1.807 (1.584)	1.869 (1.706)
Adoption × Post × Medium-term	0.959 (0.957)	0.771 (1.047)	0.593 (0.497)	0.589 (0.559)	1.479 (1.029)	2.239* (1.109)
Adoption × Post × Long-term	-0.524 (0.816)	1.269 (0.892)	0.717 (0.424)	0.362 (0.477)	0.176 (0.877)	2.010* (0.945)
Constant	4.301*** (0.152)	13.04*** (0.166)	1.649*** (0.079)	2.194*** (0.089)	5.503*** (0.164)	14.481*** (0.176)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.007	0.120	0.002	0.055	0.009	0.121

Note: Results for the voice channel are omitted as the average pre-treatment item price for adopters and non-adopters could not be observed. Robust standard errors are in parentheses. *p<0.05; **p<0.01; ***p<0.001

We then examine the treatment effects on total purchase quantity across channels over time.

In Table 9, we observe positive and significant effects on the quantity of repeat purchases in the PC channel and the effects persist across different time periods (Column 4). For the purchase quantity of new products, we observe a marginally significant quantity increase in the mobile channel. Together with the results reported in Tables 5 and 7, our findings imply that the positive spillover effects in the repeat purchase via the PC channel are driven by the increasing purchase quantity rather than a higher price. Additionally, the overall positive impact on spending across all channels is largely attributed to consumers buying repeatedly purchased products with higher prices through the voice channel.

Table 9. Treatment Effect on Total Quantity

	<i>Total Quantity</i>					
	Mobile		PC		All	
	(1) <i>New Purchase</i> (RMB)	(2) <i>Repeat Purchase</i> (RMB)	(3) <i>New Purchase</i> (RMB)	(4) <i>Repeat Purchase</i> (RMB)	(5) <i>New Purchase</i> (RMB)	(6) <i>Repeat Purchase</i> (RMB)
Adoption × Post × Short-term	0.106 (0.255)	1.669 (1.082)	0.009 (0.006)	0.045*** (0.009)	0.173 (0.255)	1.984 (1.082)
Adoption × Post × Medium-term	0.356* (0.166)	0.348 (0.703)	0.002 (0.004)	0.020*** (0.006)	0.362* (0.166)	0.431 (0.703)
Adoption × Post × Long-term	0.118 (0.141)	0.336 (0.599)	0.001 (0.003)	0.014** (0.005)	0.120 (0.141)	0.378 (0.599)
Constant	0.131*** (0.026)	1.545*** (0.112)	0.023*** (0.001)	0.043*** (0.001)	0.154*** (0.026)	1.589*** (0.112)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.001	0.147	0.011	0.202	0.001	0.148

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

There might exist an alternative explanation that voice AI may recommend users deals or promotions with higher prices, thereby driving increased monetary spending. However, this possibility can be ruled out as the Genie voice AI recommends the same set of products regardless of the channel in which consumers conduct their search (we present further details in Appendix A2). The above analyses on the total quantity and average price of items purchased further help rule out this alternative explanation. Besides, Tmall Genie will not initiate any reminders if users do not enter any search queries. Instead, users use Tmall Genie to buy items that they are familiar with. Therefore, the increased spending is not a result of any reminder effect.

9.4. Robustness Checks

To further validate our main results, we conducted a series of robustness checks, falsification tests, and additional empirical analyses.

9.4.1. Placebo Treatments Test

As the DiD approach requires the assumption of parallel pre-treatment trends, we validate this assumption by conducting a placebo treatment test. Specifically, we define a set of placebo “treatments” at three midpoints in the pre-treatment period (namely, weeks 6, 7, and 8), and we re-estimate the DiD model on the same matched sample prior to the actual treatment week. For example, if we set week 6 as the placebo treatment week, then we use weeks 1–5 as the pre-treatment period and weeks 7–13 as the post-treatment period. Theoretically, the effect of these placebo “treatments” should be insignificant, given that they are not actual interventions and no real treatment occurred during this period (i.e., before Week 13th). The empirical evidence fails to reject the null hypothesis of no treatment effect for the placebo “treatments” (see results in Appendix A7). This suggests that pretreatment trends are statistically equivalent (i.e., “parallel”) between the treatment and control groups.

9.4.2. Temporal Selection Test and Dynamic Effects

There may exist a within-consumer temporal selection issue that affects the timing of Genie adoption. If time-varying unobserved factors systematically prompt consumers to buy Genie concomitantly with other products on Tmall, then the increase in spending after adoption might not be attributable to Genie adoption. For instance, if a consumer relocates to a new house, the move may prompt them to buy a Genie. However, they may also need to buy many other products due to the move. In this case, the true but unobserved cause of their increased spending would be the move, rather than the adoption of the Genie. We address this concern by incorporating an additional interaction term to the DiD model:

$$y_{it} = \beta \cdot Post_t + \gamma \cdot (Adoption_i \times Post_t) + \rho \cdot (Adoption_i \times 1week_pre_t) + \delta_i + \tau_t + \varepsilon_{it} \quad (6)$$

where $1week_pre_t$ is an indicator of the week prior to the adoption week. If an unobserved factor is driving both adoption and increased spending, then $Adoption_i \times 1week_pre_t$ should be positive. However, the result (reported in Appendix A8) shows a negative and insignificant coefficient of $Adoption_i \times 1week_pre_t$. We conclude that within-consumer temporal selection is not a major threat to our identification.

Furthermore, we estimate a relative time model (Autor, 2003; Greenwood and Wattal, 2017), where the event is defined as the adoption of the voice AI (i.e., the fourteenth week).³⁴ As a standard procedure, the first lag (one week before the event) is set as the baseline and therefore omitted. We then estimate the following model:

³⁴ Our choice to aggregate the data into short-, mid-, and long-term periods was driven by the observation of significant fluctuations in the week-by-week results. These fluctuations could potentially make the results more difficult to interpret. We thank an anonymous reviewer for this suggestion.

$$\begin{aligned}
y_{it} = & \lambda_{LT}^{Pre} \cdot Adoption_i \times Pre_t \times LongTerm_t \\
& + \lambda_{MT}^{Pre} \cdot Adoption_i \times Pre_t \times MediumTerm_t \\
& + \lambda_{ST}^{Pre} \cdot Adoption_i \times Pre_t \times ShortTerm_t \\
& + \lambda_{ST}^{Post} \cdot Adoption_i \times Post_t \times ShortTerm_t \\
& + \lambda_{MT}^{Post} \cdot Adoption_i \times Post_t \times MediumTerm_t \\
& + \lambda_{LT}^{Post} \cdot Adoption_i \times Post_t \times LongTerm_t \\
& + \delta_i + \tau_t + \varepsilon_{it}
\end{aligned} \tag{7}$$

The estimation results in Table A8.2 present the dynamic effects over time. All the coefficients of the pre-treatment dummies are statistically indistinguishable from zero, confirming the absence of pre-trends. The magnitude and statistical significance of the post-treatment dummies decline over time, which indicates the diminishing positive effect of the Genie adoption.

9.4.3. An Alternative Aggregation Level of the Category

For the main estimation, we calculated propensity scores based on purchases in the top 13 categories plus a 14th “all others” category (aggregated from the remaining 124 categories). Though we believe that the granularity of this calculation is effective for identifying similar consumers, we conduct an alternative specification with less granularity to ensure the robustness of the findings. Specifically, we aggregate all products into just four categories: electronics, travel services, food, and all others. We repeat the matching procedure and DiD estimation. The results (reported in Appendix A9) are qualitatively the same as the main results.

9.4.4. An Alternative Measure of New vs. Repeat Purchases

Additionally, we use an alternative measure of new vs. repeat purchases. Specifically, we calculate the proportional changes defined with the average spending amount of an individual on new and repeat purchases before the adoption as the baseline.³⁵ We then re-estimate the model and report the result in Appendix A10. The analysis reveals that the treatment effect on new purchases is not

³⁵ We thank an anonymous reviewer for this suggestion.

statistically different from zero, whereas the treatment effect on repeat purchases is significantly positive, thereby confirming our main results.

9.4.5. Different Pre-Study Windows in Defining New vs. Repeat Purchases

In Section 6 and Section 8, we define new vs. repeat purchases using a pre-study window of 6 weeks. To ensure the robustness of the selection of time window, we also re-estimate all the models relying on such a definition using a pre-study window of 3, 4, 5, 7, and 8 weeks. The results are all qualitatively similar as shown in Appendix A11.

9.4.6. An Additional Dataset and the IV Method

As described in section 3.1, a potential concern with our empirical setting is that consumers may choose to adopt Genie due to some nonrandom shocks that are unobservable (and thus cannot be incorporated into our matching procedure) and also cannot be absorbed by the individual and time fixed effects. In Appendix A13, we provide extra empirical evidence using an additional dataset collected from Alibaba in a different time period; we take advantage of an essentially random coupon allocation event to use an instrumental variable approach. The estimated main effect is qualitatively consistent with the effect presented in the main text.

10. Implications and Conclusions

Voice shopping holds increasing significance in our daily lives and the online economy. This paper takes the first step toward understanding the impact of consumer adoption of voice shopping technology on subsequent spending within the affiliated e-commerce platform. We employ granular propensity score matching to match adopters with non-adopters and thereafter use the difference-in-differences approach to identify the effect of Genie adoption on spending.

Our analyses yield the following main findings: *First*, our findings indicate that the adoption of Tmall Genie, the voice AI device launched by e-commerce giant Alibaba, leads consumers in our sample to increase their average weekly spending on Tmall by 16.6% within the first four months after adoption. *Second*, Genie adoption has stronger positive effects on spending for repeat purchases, less substitutable and more familiar products, all of which do not require consumers to digest product information at high costs. Constrained by the inconvenience of searching and comparing products using voice, consumers prefer to purchase products that do not require complicated selection processes through the voice channel. *Third*, while the positive effect of Genie adoption on spending attenuates over time, it continues to be significantly positive in the long term, especially for repeat purchases in the voice channel. Because of the novelty effect and the shortcomings of information processing using voice, the long-term positive effect is mainly attributed to repeat purchases. *Lastly*, Genie adoption produces a positive and significant spillover effect on spending via the PC channel, while its impact on mobile channel spending is insignificant. The category-level analysis illustrates that the direction of the spillover effect on the two channels, whether positive or negative, depends on the specific shopping context.

Our study provides several managerial implications for the development of voice AI technology in e-commerce platforms. *First*, we demonstrate that the adoption of a voice AI device facilitates additional online spending, so voice AI technologies may be well-positioned to enhance the revenue growth of e-commerce platforms. When designed well, voice AI devices as a new form of consumer-platform interaction can improve user experience and help companies build winning relationships with their customers,³⁶ but there can be challenges in terms of information processing that limit its potential.

³⁶ Cited from press: <https://www.forbes.com/sites/forbestechcouncil/2021/08/24/harnessing-conversational-voice-ai-in-the-e-commerce-industry/?sh=432677f060ec>.

Our findings show that the positive effects of voice AI adoption on spending are mainly attributed to the product categories that do not require consumers to do extensive search or comparison, namely, repeat purchases, and more familiar and less substitutable product categories. To realize the fuller potential of voice AI technology, future product developers could take this and be cognizant of the cost of information processing in their product design. For instance, future designs could include features such as small screen displays on the Genie devices to complement the shortcomings of voice AI, such as unclear or unnatural representation of information. This can help decrease consumers' perceived amount of effort needed to interact with voice AI, and therefore adopt voice shopping for a wider range of products. Another way is to present choices and highlight information in a manner that consumers can easily understand and evaluate through speech. For example, companies can leverage the purchase histories of Genie users and personalize product recommendations for repeat purchases and more familiar products via the voice marketing channel. Such targeted marketing practices could potentially lead to reduced marketing expenditures and a higher return on investment for managers.

Second, our results show that the adoption of voice shopping has a positive spillover effect on spending in the PC channel (and no significant effect on spending in the mobile channel). One possible cause of the complementary relationship is that consumers may combine the voice channel with another channel to complete a purchase. Managers may want to keep this positive spillover effect in mind when designing e-commerce platforms for optimize customer engagement. *Moreover*, our study finds that engagement and motivation for continued usage of voice AI decline over time due to the novelty effect. Future product designs can focus on creating more consumer-oriented technology, which can hopefully maintain consumers' interest in using voice AI. For instance, companies can adjust the emotional-expressing capabilities (Han et al., 2022) and

incorporate human-like behavior into AI agents (Schanke et al., 2021) to improve their performance and efficacy, so that even after the short-term curiosity disappears, the higher perceived usefulness and better human-computer interaction can remain attractive to consumers in the long run. *Lastly*, our research extends its implications beyond the realm of e-commerce to other business sectors. For example, in the hospitality industry, Marriott hotels now include Amazon Alexa; guests can order room service by talking to a voice-activated ordering device in their rooms. The heterogeneity and moderating effects we document in the effects of voice-activated shopping devices may transfer to offline voice-aided purchases (or demand for services) in the context of hotels.

We conclude by identifying several questions beyond this paper's scope, yet they represent potential avenues for future research. First, we are unable to access the specific product information, but only category information, for each transaction. More granular data on consumers' purchase journeys and transaction details, particularly the search aspect, would have allowed for a deeper exploration of the heterogeneous effects and the nuanced mechanisms between channels. Second, as consumers become more familiar with AI products, the impact of voice AI device adoption on shopping behaviors may diminish. Our findings on the short-term and long-term effects partially address this issue, but we urge caution in extrapolating the effect to an even longer-term (i.e., beyond 17 weeks post-adoption). Third, though we have leveraged several econometric techniques to address the potential endogeneity concerns, more can be done to more accurately identify the impact of Genie adoption on a consumer's Tmall spending, such as analyses on adopters who received the device without self-selection. Moreover, to provide more evidence on the underlying mechanism behind the channel spillover effects, future researchers could exploit more granular data on consumers' cross-channel purchase journeys, which we are unfortunately

unable to collect and process. Lastly, in this study, we are agnostic about how voice AI technology can better serve consumers, but we hope future research can explore this question with richer data of full conversations between voice AI agents and consumers.

References

- Adamopoulos P, Todri V, Ghose A. (2020). Demand effects of the internet-of-things sales channel: Evidence from automating the purchase process. *Information Systems Research*, 32(1), 238-267.
- Alexander RS. (1948). Report of the definitions committee. *Journal of Marketing* (pre-1986), 13(000002), 202.
- Autor DH. (2003). Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21(1), 1-42.
- Ba W, Mendelson H, Zhu M (2020) Sales Policies for a Virtual Assistant. ArXiv preprint arXiv: 2009.03719.
- Bahmani N, Bhatnagar A, Gauri D. (2022). Hey, Alexa! What attributes of Skills affect firm value? *Journal of the Academy of Marketing Science*, 50(6), 1219-1235.
- Bergner AS, Hildebrand C, Häubl G. (2023). Machine Talk: How Verbal Embodiment in Conversational AI Shapes Consumer–Brand Relationships. *Journal of Consumer Research*, forthcoming.
- Bronnenberg B, Dubé JP, Mela C (2010) Do digital video recorders influence sales? *Journal of Marketing Research* 47(6):998–1010.
- Brynjolfsson E, Li D, Raymond LR. (2023). Generative AI at work (No. w31161). *NBER Working Paper*.
- Chen G, Xiao S, Zhang C, Zhao H. (2023). A theory-driven deep learning method for voice chat–based customer response prediction. *Information Systems Research*, 34(4), 1513-1532.
- Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1(2), 203-222.
- Datta H, Knox G, Bronnenberg BJ (2018) Changing their tune: How consumers’ adoption of online streaming affects music consumption and discovery. *Marketing Science*, 37(1), 5–21.
- Forbes (2023). Reshaping E-Commerce: The Influence Of AI-Generated Content. Accessed October 3, 2023, <https://www.forbes.com/sites/shoplazza/2023/08/02/reshaping-e-commerce-the-influence-of-ai-generated-content/>.
- Ghose A (2018) TAP: Unlocking the mobile economy. MIT Press.
- Ghose A, Goldfarb A, Han SP (2013) How is the mobile Internet different? Search costs and local activities. *Information Systems Research*, 24(3), 613-631.
- Greenwood BN, Wattal S. (2017). Show me the way to go home. *MIS Quarterly*, 41(1), 163-188.

- Goli A, Singh A. (2024). Frontiers: Can Large Language Models Capture Human Preferences? *Marketing Science*.
- Han E, Yin D, Zhang H. (2022). Bots with Feelings: Should AI Agents Express Positive Emotion in Customer Service. *Information Systems Research*.
- Huang MH. (2003). Designing website attributes to induce experiential encounters. *Computers in Human Behavior*, 19(4), 425-442.
- Imbens GW, Rubin DB (2015) Causal Inference in Statistics, Social, and Biomedical Sciences (Cambridge University Press, Cambridge, UK).
- Kleijnen M, De Ruyter K, Wetzels M. (2007). An assessment of value creation in mobile service delivery and the moderating role of time consciousness. *Journal of Retailing*, 83(1), 33-46.
- Kumar V, Dixit A, Javalgi RRG, Dass M (2016) Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing. *Journal of the Academy of Marketing Science*, 44(1), 24-45.
- Lee RS (2013) Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7), 2960-3000.
- Li P, Castelo N, Katona Z, Sarvary M. (2024). Frontiers: Determining the Validity of Large Language Models for Automated Perceptual Analysis. *Marketing Science*.
- Luo X, Qin MS, Fang Z, Qu Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing*, 85(2), 14-32.
- 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.
- Manchanda P, Packard G, Pattabhiramaiah A. (2015). Social dollars: The economic impact of customer participation in a firm-sponsored online customer community. *Marketing Science*, 34(3), 367-387.
- Massetti B. (1996). An empirical examination of the value of creativity support systems on idea generation. *MIS Quarterly*, 83-97.
- McAfee RP, McMillan J (1987). Auctions and bidding. *Journal of Economic Literature*, 25(2), 699-738.
- McMillan R (2007) Different Flavor, Same Price: The Puzzle of Uniform Pricing for Differentiated Products, mimeo, Federal Trade Commission
- Melumad S. (2023). Vocalizing search: How voice technologies alter consumer search processes and satisfaction. *Journal of Consumer Research*, 50(3), 533-553.
- Meuter ML, Ostrom AL, Roundtree RI, Bitner MJ. (2000). Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, 64(3), 50-64.
- Munz KP, Morwitz VG. (2022). Sound Judgment: Evaluability and Memory in Speech-based Product Evaluation and Choice. *Working Paper*.
- Pamuru V, Khern-am-nuai W, Kannan K. (2021). The impact of an augmented-reality game on local businesses: A study of Pokémon Go on restaurants. *Information Systems Research*, 32(3), 950-966.

- Park Y, Bang Y, Ahn JH. (2020). How does the mobile channel reshape the sales distribution in E-commerce? *Information Systems Research*, 31(4), 1164-1182.
- Parthasarathy M, Bhattacharjee A. (1998). Understanding post-adoption behavior in the context of online services. *Information Systems Research*, 9(4), 362-379.
- Schanke S, Burtch G, Ray G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736-751.
- Shiller BJ, Waldfogel (2009) Music for a Song: An Empirical Look at Uniform Song Pricing and its Alternatives, NBER Working Paper: 1539
- Smith KT (2020) Marketing via smart speakers: what should Alexa say? *Journal of Strategic Marketing*, 28(4), 350-365.
- Son Y, Oh W, Im I. (2023). The Voice of Commerce: How Smart Speakers Reshape Digital Content Consumption and Preference. *MIS Quarterly*, 47(2), 857-874.
- Sun C, Adamopoulos P, Ghose A, Luo X. (2022). Predicting stages in omnichannel path to purchase: A deep learning model. *Information Systems Research*, 33(2), 429-445.
- Tan TF, Netessine S (2020) At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science*, 66(10), 4496-4515.
- Wooldridge J. (2010). *Econometric Analysis of Cross Section and Panel Data*. Econometric Analysis of Cross Section and Panel Data. MIT Press
- Xu J, Forman C, Kim JB, Van Ittersum K (2014) News media channels: Complements or substitutes? Evidence from mobile phone usage. *Journal of Marketing*, 78(4), 97–112.
- Xu K, Chan J, Ghose A, Han SP (2017) Battle of the channels: The impact of tablets on digital commerce. *Management Science*, 63(5), 1469-1492.
- Xu Y, Ghose A, Xiao B. (2023). Mobile Payment Adoption: An Empirical Investigation of Alipay. *Information Systems Research*, forthcoming.
- Zhou M, Geng D, Abhishek V, Li B. (2020). When the Bank comes to you: Branch network and customer Omnichannel banking behavior. *Information Systems Research*, 31(1), 176-197.

Appendices

A1. Features of Popular Voice Devices

We investigate Alibaba Tmall Genie, which is the most popular voice AI device in China. Tmall Genie has been marketed as a shopping device, but it offers most of the same functions as Amazon Echo, including smart home control, weather, news, and music.³⁷ The consumer begins the activation of all features by saying “Tmall Genie” in Mandarin. As the name suggests, consumers can use Tmall Genie to place verbal orders for items from Tmall, Alibaba’s shopping site; Genie has voiceprint recognition to ensure that each order comes from an authorized user. The table³⁸ below compares the features of four popular voice AI devices.

	Tmall Genie	Amazon Echo	Google Home	Apple Homepod
Shopping	✓	✓	✓	×
Track packages	✓	✓	×	×
Web search	✓	✓	✓	✓
Smart home control ³⁹	✓	✓	✓	✓
News/Music	✓	✓	✓	✓
Phone call	✓	✓	✓	✓
Skills ⁴⁰	✓	✓	✓	✓
Read audiobook	✓	✓	✓	×
Intercom ⁴¹	×	×	✓	×

³⁷ Cited from: <https://www.theverge.com/circuitbreaker/2017/7/5/15921722/alibaba-amazon-echo-competitor-tmall-genie-beta-voice-speaker-shop>.

³⁸ The unique features of each device can be found here: <https://thewiredshopper.com/google-home-vs-amazon-echo-vs-apple-homepod/#tab-con-11>. We also consulted other online resources to create the table.

³⁹ Some smart home devices are compatible with only some of the smart speakers. For example, thermostats and security cameras from Nest, as well as Samsung SmartThings, are compatible with only Amazon Echo and Google Home.

⁴⁰ Skills are called “Actions” for Google Home and “Abilities” for Apple Homepod.

⁴¹ Cited from: <https://www.forbes.com/sites/forbes-personal-shopper/2019/05/14/apple-homepod-vs-amazon-echo-vs-google-home-which-smart-speaker-is-right-for-you/#15dd19df10ef>.

A2. Institutional Details of the Voice Shopping Process

We provide two important institutional details about the use of Tmall Genie for shopping. First, Genie recommends products (in addition to providing a search function). Alibaba currently uses the same ranking algorithm to sort products based on the consumer's query, no matter if the query comes from the web interface or from Genie. This means the products recommended by Genie after a consumer's inquiry is no different from those that would have been recommended on other channels (namely, PC and mobile). It is therefore unlikely that the voice AI may intentionally recommend products of large size and higher prices. Below is a representative conversation.

<i>Customer</i>	“Tmall Genie, I want to buy potato chips.”
<i>Genie</i>	“Master, I recommend Dali Park Kebby potato chips four packs, each pack 32 grams, free shipping, actual price 9.98 yuan. It will be delivered to (address), may I place the order for you?”
<i>Customer</i>	“Yes, please place the order.”
<i>Genie</i>	“Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the ‘beep,’ say ‘Tmall Genie, 9951.’” (Here, 9951 is the authentication code randomly generated by the system.)
<i>Customer</i>	“9951.”
<i>Genie</i>	“Alipay discount is applied. If you want to know the delivery status, you can let me know by saying ‘Tmall Genie, tracking information.’”

The consumer may also specify the desired brand:

<i>Customer</i>	“Tmall Genie, I want to buy some Lay's potato chips.”
<i>Genie</i>	“Master, I recommend Lay's three-piece canned snack food package, shipping cost 20 yuan, actual price 37.9 yuan. It will be delivered to (address), may I place the order for you?”
<i>Customer</i>	“Not this time.”
<i>Genie</i>	“How about Lay's black pepper cheese non-fried potato chips? Shipping cost 20 yuan, actual price 37 yuan.”
<i>Customer</i>	“No thanks!”
<i>Genie</i>	“Ok, looking forward to your next use.”

In both examples, Genie provides only one product recommendation in each answer, and the recommended product is the same as the first result listed on the Tmall website in response to the equivalent text-based search. Specifically, In the first example, when a consumer searches for “potato chips” on the Tmall website, the first returned result is “Dali Park Kebby potato chips four

packs, each pack 32 grams, free shipping, actual price 9.98 yuan”; in the second example, when a consumer enters “Lay’s potato chips,” the first returned result is “Lay’s three-piece canned snack food package, shipping cost 20 yuan, actual price 37.9 yuan,” and the second result is “Lay’s black pepper cheese non-fried potato chips, shipping cost 20 yuan, actual price 37 yuan.”

Second, if the consumer does not fully specify the product (e.g., “order a refill card” without specifying the amount), Genie recommends a product with a specific payment amount. The Genie voice AI recommends the same set of products regardless of the channel in which consumers conduct their search, and will not proactively initiate any conversation if users did not enter any search queries.

For example:

<i>Customer</i>	“Tmall Genie, I’d like to order a mobile refill card.”
<i>Genie</i>	“Master, I would recommend China Mobile’s refill card. The total price is 100 RMB. It will be delivered to (address). May I place the order for you?”
<i>Customer</i>	“Yes, please place the order.”
<i>Genie</i>	“Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the ‘beep,’ say ‘Tmall Genie, 2065.’” (Here, 2065 is the authentication code randomly generated by the system.)
<i>Customer</i>	“2065.”
<i>Genie</i>	“Alipay discount is applied. If you want to know the delivery status, you can let me know by saying ‘Tmall Genie, tracking information.’”

As in the first two examples, the recommended product is the first returned result when the search query, “refill card,” is entered on the Tmall website.

A3. Adoption Rate across Demographics

Figure A3. Adoption Rate by Customer Age and Gender

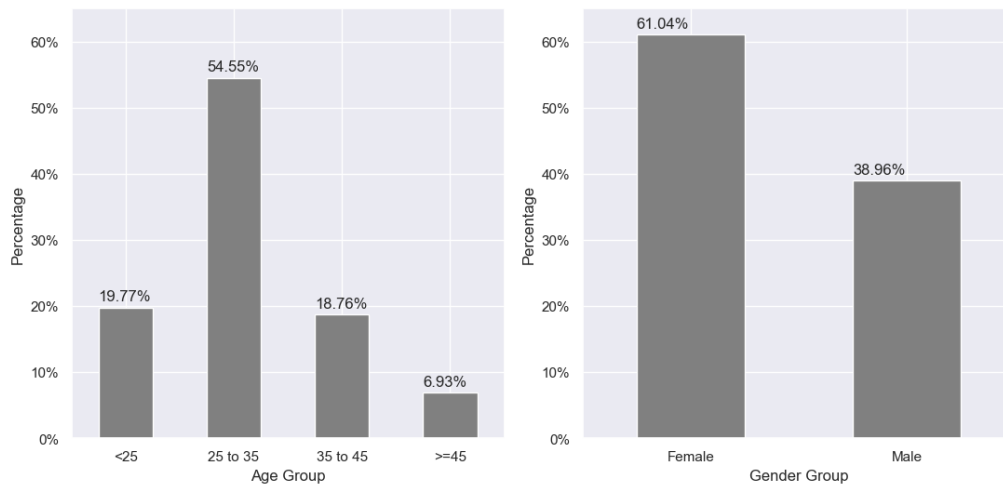


Table A3. Bayesian Posterior Adoption Rate

	Observation	Prior	Posterior
Adoption		2.15%	
Female	61%	58.64%	2.24%
Male	39%	41.36%	2.03%
Age < 25	19.8%	43.75%	0.97%
Age in [25, 35)	54.6%	44.39%	2.64%
Age in [35, 45)	18.8%	7.66%	5.28%
Age \geq 45	6.9%	4.2%	3.53%

Note: We employ a Bayesian rule to derive the posterior estimation of the adoption rate across genders and age groups. For example, $\Pr(\text{Adoption} | \text{Male}) = \Pr(\text{Male} | \text{Adoption}) * \Pr(\text{Adoption}) / \Pr(\text{Male})$. The observations in column (1) come from our data, whereas the priors in column (2) are from consulting firms. For example, some sample statistics can be found at <https://n.znds.com/article/37815.html> and <https://www.iimedia.cn/c1061/69035.html>. The posterior probability can inform us about the adoption rate in a more intuitive way than the observed likelihood. For example, to determine the adoption rate of males vs. females, one needs to report $\Pr(\text{Adoption} | \text{Male}) = 2.23\%$ instead of $\Pr(\text{Male} | \text{Adoption}) = 39\%$. The last column of this table can be used to answer questions of this kind.

Table A3 shows that 2.24% of female consumers and 2.03% of male consumers in our sample adopted Genie. Consumers in the age group 35–45 have the highest adoption rate (5.28%), while consumers younger than 25 have the lowest adoption rate (0.97%).

A4. List of All 137 Product Categories⁴²

Agriculture Equipment	Entertainment Gift Card	O2O Store
Anime	Farm & Agricultural Supplies	Office & School Supplies
Art & Collectibles	Fashion Accessories	Office Equipment
Auction	Flower Services	Office Furniture
Audio Visual Electronics	Food Ware	OTC Medicines
Baby Diapers	Fresh Hema	Others
Baby Dietary Supplement	Furniture	Packaging
Baby Products	Game Time Cards	Personal care
Bags	Gift Cards (Offline)	Personalized Interior Design
Bath & Beauty	Government Assets	Pet Supplies
Bedding	Graphic Design Services	Pocket Knives & Glasses
Bicycles & Accessories	Hair Care Products	Pregnancy & Maternity Products
Books	Hardware & Tools	Prescription Drug
Building Material	Health & Fitness	Restaurant Gift Cards
Cainiao Shop	Holiday Shop	Retail Logistics Services
Cameras	Home & Living	Sex Toys
Car Care & Electronics	Home Appliances	Shopping Gift Cards
Car Maintenance Services	Home Carpet & Flooring	Smart Home
Cars	Home Décor Fabric	Snacks
Cash Cards	Home Decoration	Software
Charity	Hotel	Spa
Child Care Supplies	House Painting & décor	Spa Supplies
Children's Shoes	In-Game Item Marketplace	Sport shoes
Clean Energy Transportation	Insurance	Sporting Goods
Cleaning Supplies	Jewelry	Sports & Outdoors
Coffee & Beverages	Jewelry & Accessories	Sports Accessories
Commercial Appliances	Kids' Clothing	Sports Clothing
Computers & accessories	Kitchen Appliances	Storage & Organizer
Contact lenses	Kitchen Supplies	Store Coupons
Cosmetics	Laptops	Tablets
Coupons	Latex Products	Taobao Gift Card
Craft Supplies	Lights	Tea
Custom Design	Local Services	Tencent QQ
Data Storage	Loungewear	Tmall O2O retailer
Desktop Computers	Loyalty Card	Tours & Activities
Dietary Supplements	Meat & Vegetables	Toys
Digital Reading	Medical Services	Transportation
DIY PC	Medication Delivery Services	Travel Services
Donation	Men's Clothing	Used Electronics
Door-to-door Services	Men's Shoes	Used Good
Dry Goods & Seasoning	Mobile Phone Plans	Video Game
Education & Training	Mobile Phones	Watches
Ele.me Delivery Services	Mobile Refill Card	Wellness
Electric Vehicle	Motorcycles	Wine & Liquor
Electrical Engineering	Movies & Sports Ticket	Women's Clothing
Electronic Components	Music & Movies	Women's Shoes
Electronics	Music Instruments	
Electronics Accessories	Network Devices	

⁴² *Note:* In the mobile, PC, and voice channels, consumers spend the most on the *Mobile Refill Card*, *Movies & Sports Tickets*, and *Mobile Refill Card* categories, respectively. Regrettably, due to an NDA with our data sponsor, we cannot divulge comprehensive details regarding the primary product categories in each channel.

A5. Heterogeneity across Age and Gender

Theoretically, the moderating effect of age presents ambiguity. While younger individuals may be more receptive to new technology, potentially altering their shopping behaviors, their typically greater income constraints could limit their spending flexibility. Conversely, older consumers, possibly spending more time at home and having more opportunities to interact with Tmall Genie, may encounter difficulties in adapting to new consumption habits and prefer familiar traditional methods of shopping. We categorize users into four groups: younger than 25 years old, between 25 and 35, between 35 and 45, and older than 45, and empirically estimate the moderating effect of age as follows:

$$\begin{aligned}
 y_{it} = & \lambda_{age1} \cdot (Adoption_i \times Post_t \times I[Age_i < 25]) + \\
 & \lambda_{age2} \cdot (Adoption_i \times Post_t \times I[25 \leq Age_i < 35]) + \\
 & \lambda_{age3} \cdot (Adoption_i \times Post_t \times I[35 \leq Age_i < 45]) + \\
 & \mu_{age1} \cdot (Post_t \times I[Age_i < 25]) + \\
 & \mu_{age2} \cdot (Post_t \times I[25 \leq Age_i < 35]) + \\
 & \mu_{age3} \cdot (Post_t \times I[35 \leq Age_i < 45]) + \\
 & \rho \cdot (Adoption_i \times Post_t) + \delta_i + \tau_t + \varepsilon_{it}
 \end{aligned} \tag{A5.1}$$

where y_{it} is the total amount spent by consumer i in week t ; Age_i is the age of consumer i ; the indicator variables (e.g., $I[Age_i < 25]$) are set to 1 if the consumer's age falls in the specified range; and λ captures the difference between the moderating effect of the given age range and of the reference age group. Note that the age group $I[Age_i \geq 45]$ is selected as the baseline reference group and the term $Post$ is absorbed by the time fixed effects. In a similar vein, we estimate the moderating effect of gender as follows:

$$\begin{aligned}
 y_{it} = & \lambda_{female} \cdot (Adoption_i \times Post_t \times I[Gender_i = female]) \\
 & \mu_{female} \cdot (Post_t \times I[Gender_i = female]) \\
 & + \rho \cdot (Adoption_i \times Post_t) \\
 & + \delta_i + \tau_t + \varepsilon_{it}
 \end{aligned} \tag{A5.2}$$

where y_{it} is the total amount spent by consumer i in week t ; $I[Gender_i = female]$ is set to 1 to indicate consumer i 's gender being female; and λ_{female} captures the difference between the moderating effect of two gender groups. Note that the gender group $I[Gender_i = male]$ is selected as the baseline reference group and the term $Post$ is absorbed by the time fixed effects.

Table A5. Heterogeneity in the Treatment Effect across User Demographics

	(1)	(2)
	<i>Spending (RMB)</i>	<i>Spending (RMB)</i>
Adoption \times Post	-20.85** (8.060)	6.833* (3.220)
Adoption \times Post \times Age < 25	31.05*** (9.269)	
Adoption \times Post \times Age in [25, 35)	27.66*** (8.510)	
Adoption \times Post \times Age in [35, 45)	29.06*** (9.272)	
Post \times Age < 25	-0.4489 (4.550)	
Post \times Age in [25, 35)	0.436 (4.242)	
Post \times Age in [35, 45)	2.258 (4.455)	
Adoption \times Post \times Female		-1.847 (4.132)
Post \times Female		0.685 (1.734)
Constant	32.53*** (2.359)	32.69*** (0.738)
Individual FEs	Yes	Yes
Week FEs	Yes	Yes
Observations	120,435	120,435
Adjusted R ²	0.109	0.109

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table A5 reports the estimated moderation effects of the two consumer demographics, age (Column 1) and gender (Column 2), as specified above. The results indicate that the positive effect of Genie adoption on spending is the strongest for users younger than 25 (+ RMB 10.2 weekly, $p < 0.001$), followed by users in the age group [35, 45) (+ RMB 8.21 weekly, $p < 0.001$) and the age group [25, 35) (+ RMB 6.81 weekly, $p < 0.001$). Interestingly, the effect is *negative* for those older

than 45 (- RMB 20.85 weekly, $p < 0.01$). A possible explanation is that older consumers have a harder time learning to use the voice shopping assistant, and the increased difficulty of the task deters online shopping. For gender, we find that the positive effect of Genie adoption on purchase spending is directionally yet insignificantly stronger for male consumers than for female consumers (+ RMB 6.833 vs. + RMB 4.986, $p > 0.1$). In summary, although Genie adoption has a significant and positive effect on spending for most consumer demographics, managers and marketers should expect different scales of impacts on spending by considering the moderating effects of age and gender.

A6. Treatment Effect Across Channels for Elder Consumers**Table A6. Treatment Effect Across Channels for AGE ≥ 45**

	(1) <i>Mobile Spending</i> (RMB)	(2) <i>PC Spending</i> (RMB)		(3) <i>Voice Spending</i> ⁴³ (RMB)
Adoption \times Post	-24.70*** (0.041)	-2.185 (1.539)	Post	3.370** (0.630)
Constant	25.80*** (0.009)	3.851*** (0.320)	Constant	0.000 (0.480)
Individual FEs	Yes	Yes	Individual FEs	Yes
Week FEs	Yes	Yes	Week FEs	No
Observations	8,928	8,928	Observations	1,488
Adjusted R ²	0.121	0.062	Adjusted R ²	0.016

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

We derive the following conclusions from Table A6. First, in the Voice channel, we find positive and significant effects. Second, in the PC channel, we observe an insignificant spillover effect on spending for middle-aged consumer subgroups. For these adopters, Genie adoption does not have a complementary effect spilling over to the PC channel, which is distinct from the respective overall spillover effect (Table 7). Further, it is noteworthy that there are negative and significant spillover effects on the spending in the mobile channel. One possibility is that older consumers spend more time at home than younger consumers and therefore have less need for a portable or mobile shopping channel. In addition, it can be difficult for them to effectively navigate mobile devices that are inadequate due to the tiny screens, keyboards, and limited communication speeds. As an alternative, voice AI can be seen as a more user-friendly and convenient option as it only requires the use of spoken commands. This preference for voice-based interaction over mobile devices may have resulted in middle-aged adopters making fewer purchases through the mobile channel.

⁴³ The model specification for column (3) is the same as Equation (2) except that Adoption = 1 and Week FEs are not included.

A7. Placebo Test Results

Table A7. Main Effects using Placebo Treatments

Variable	Placebo Tr. at Week 6 (RMB)	Placebo Tr. at Week 7 (RMB)	Placebo Tr. at Week 8 (RMB)
Adoption \times Post	-0.720 (3.121)	0.567 (3.121)	-1.711 (3.198)
Constant	33.78*** (0.667)	33.66*** (0.649)	33.82*** (0.635)
Individual FEs	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes
Observations	50,505	50,505	50,505
Adjusted R ²	0.112	0.112	0.112

Note: Robust standard errors are in parentheses.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

A8. Temporal Selection Robustness Test Results

Table A8.1. Main Effect with an Additional Interaction Term Included

	Spending (RMB)
Adoption × Post	4.870* (2.066)
Adoption × 1Week_Pre	-11.04 (5.770)
Constant	33.07*** (0.442)
Individual FEs	Yes
Week FEs	Yes
Observations	120,435
Adjusted R ²	0.109

Note: Robust standard errors are in parentheses.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A8.2. Relative Time Model and Dynamic Effects

Variables	Total Spending (RMB)	
	Estimates	SE
Pretreatment Trends:		
Adoption × Pre × Long-term	3.155	(5.105)
Adoption × Pre × Medium-term	4.894	(4.914)
Adoption × Pre × Short-term	-7.421	(6.641)
Post-treatment Trends:		
Adoption × Post × Short-term	12.93**	(4.971)
Adoption × Post × Medium-term	10.56*	(4.464)
Adoption × Post × Long-term	6.350	(4.395)
Individual FEs	Yes	
Week FEs	Yes	
Observations	120,435	
Adjusted R ²	0.109	

Note: Robust standard errors are in parentheses. The week before treatment is set as the baseline.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001.

A9. Robustness Check Results using Aggregated Product Categories**Table A9.1. Summary Statistics with Aggregated Product Categories**

Variable	N	Mean	SD	Min.	Max.
User Demographics					
<i>Age</i>	38,179	31.89	7.656	18	57
<i>Gender (male = 1)</i>	38,179	0.583	0.493	0	1
<i>Number of Historical Purchase</i>	38,179	301.2	461.8	0	13,641
Aggregated Categories					
<i>Spending_Electronics</i>	1,183,549	11.77	47.44	0	24,000
<i>Spending_Travel Services</i>	1,183,549	4.559	60.43	0	9,368
<i>Spending_All others</i>	1,183,549	4.411	67.06	0	24,294
<i>Spending_Food</i>	1,183,549	0.062	5.731	0	3,600
<i>Browsing_SmartDevices</i>	1,183,549	0.080	2.474	0	832

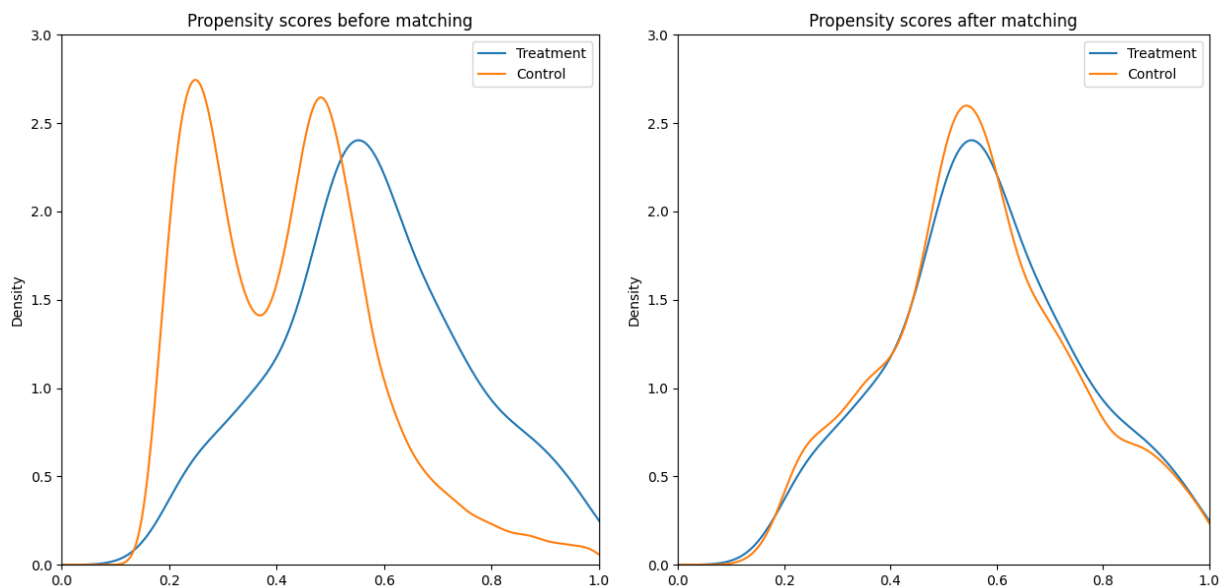
Figure A9.1. Distribution of Propensity Scores Before and After Matching with Aggregated Product Categories

Table A9.2. Comparison of Adopters and Non-Adopters Before and After Matching with Aggregated Product Categories

	Adopter (N)	Adopter (Mean)	Non- adopter (N)	Non- adopter (Mean)	p-value of t-test
(a) Before matching					
<i>Age</i>	693	31.00	37,486	31.90	0.002
<i>Gender (Male = 1)</i>	693	0.390	37,486	0.587	0.000
<i># Historical Purchases</i>	693	560.0	37,486	296.4	0.000
<i>Spending_Electronics</i>	693	17.46	37,486	11.78	0.000
<i>Spending_All others</i>	693	10.35	37,486	4.587	0.000
<i>Spending_Travel Services</i>	693	6.182	37,486	2.931	0.000
<i>Spending_Food</i>	693	0.036	37,486	0.053	0.707
<i>Browsing_SmartDevices</i>	693	0.324	37,486	0.076	0.000
(b) After matching					
<i>Age</i>	693	31.00	3,146	30.84	0.586
<i>Gender (Male = 1)</i>	693	0.390	3,146	0.423	0.106
<i># Historical Purchases</i>	693	560.0	3,146	571.7	0.702
<i>Spending_Electronics</i>	693	17.46	3,146	17.44	0.983
<i>Spending_All others</i>	693	10.35	3,146	9.154	0.536
<i>Spending_Travel Services</i>	693	6.182	3,146	6.721	0.648
<i>Spending_Food</i>	693	0.036	3,146	0.065	0.343
<i>Browsing_SmartDevices</i>	693	0.324	3,146	0.215	0.111

Table A9.3. Main Effect, Estimated on the Sample Matched with Aggregated Product Categories

	<i>All</i> <i>Spending (RMB)</i>	<i>Mobile</i> <i>Spending (RMB)</i>	<i>PC</i> <i>Spending (RMB)</i>
Adoption × Post	6.682** (2.663)	2.062 (2.550)	1.849** (0.698)
Constant	32.15*** (0.279)	27.619*** (0.267)	4.451*** (0.073)
Individual FEs	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes
Observations	119,009	119,009	119,009
Adjusted R ²	0.089	0.081	0.052

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table A9.4. Moderating Role of New vs Repeat Purchases on Main Effect, Estimated on the Sample Matched with Aggregated Product Categories

	<i>All Channels (RMB)</i>			<i>Voice Channel (RMB)</i>	
	<i>New Purchase</i>	<i>Repeat Purchase</i>		<i>New Purchase</i>	<i>Repeat Purchase</i>
Adoption × Post	0.201 (2.120)	6.482*** (1.844)	Post	0.412*** (0.048)	2.783*** (0.167)
Constant	12.15*** (0.222)	20.01*** (0.193)	Constant	0.000 (0.028)	0.000 (0.097)
Individual FEs	Yes	Yes	Individual FEs	Yes	Yes
Week FEs	Yes	Yes	Week FEs	No	No
Observations	119,009	119,009	Observations	21,483	21,483
Adjusted R ²	0.033	0.033	Adjusted R ²	0.022	0.031

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table A9.5. Heterogeneity across Product Category Features, Estimated on the Sample Matched with Aggregated Product Categories

	<i>Voice Channel</i>		
	<i>Spending (RMB)</i>	<i>Spending (RMB)</i>	<i>Spending (RMB)</i>
Post	0.000 (0.001)	0.015*** (0.001)	-0.002*** (0.001)
Post × Low Substitutability	0.823*** (0.008)		0.654*** (0.008)
Post × Familiarity		0.259*** (0.001)	0.202*** (0.001)
Constant	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Individual FEs	Yes	Yes	Yes
Category FEs	Yes	Yes	Yes
Observations	2,943,171	2,943,171	2,943,171
Adjusted R ²	0.026	0.026	0.028

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table A9.6. Novelty Effect, Estimated on the Sample Matched with Aggregated Product Categories

	(1) All <i>Total Spending</i> (RMB)		(2) Voice <i>New Purchase</i> (RMB)	(3) Voice <i>Repeat Purchase</i> (RMB)
Adoption × Post × Short-term	10.59** (3.572)	Post × Short-term	2.263*** (0.355)	8.676*** (0.599)
Adoption × Post × Medium-term	8.844** (3.231)	Post × Medium-term	0.252*** (0.068)	2.893*** (0.219)
Adoption × Post × Long-term	4.603 (3.017)	Post × Long-term	0.137*** (0.037)	1.538*** (0.207)
Constant	32.15*** (0.279)	Constant	0.000 (0.028)	0.000 (0.097)
Individual FEs	Yes	Individual FEs	Yes	Yes
Week FEs	Yes	Week FEs	No	No
Observations	119,009	Observations	21,483	21,483
Adjusted R ²	0.089	Adjusted R ²	0.013	0.052

Note: Robust standard errors are in parentheses.

*p<0.05; **p<0.01; ***p<0.001

Table A9.7. Heterogeneity in the Treatment Effect across Channels, Estimated on the Sample Matched with Aggregated Product Categories

	Mobile		PC	
	(1) <i>New Purchase</i> (RMB)	(2) <i>Repeat Purchase</i> (RMB)	(3) <i>New Purchase</i> (RMB)	(4) <i>Repeat Purchase</i> (RMB)
Adoption × Post × Short-term	0.222 (2.830)	-2.128 (2.999)	0.208 (1.167)	1.787 (1.030)
Adoption × Post × Medium-term	1.072 (1.839)	2.318 (1.949)	0.144 (0.758)	2.589*** (0.669)
Adoption × Post × Long-term	-0.473 (1.567)	2.533 (1.661)	-0.240 (0.646)	1.529** (0.570)
Constant	9.846*** (0.294)	17.77*** (0.312)	2.251*** (0.121)	2.200*** (0.107)
Individual FEs	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes
Observations	119,009	119,009	119,009	119,009
Adjusted R ²	0.028	0.088	0.013	0.061

Note: Robust standard errors are in parentheses.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

A10. Alternative Measures of New vs. Repeat Purchases**Table A10. Moderating Role of New vs Repeat Purchases on Main Effect Using an Alternative Measure**

	<i>All Channels</i>	
	<i>New Purchase</i> Proportional Change	<i>Repeat Purchase</i> Proportional Change
Adoption × Post	-7.831 (8.621)	36.745*** (13.978)
Constant	5.006*** (1.856)	3.068*** (3.002)
Individual FEs	Yes	Yes
Week FEs	Yes	Yes
Observations	119,845	118,732
Adjusted R ²	0.014	0.020

Note: Robust standard errors are in parentheses. Proportional changes are defined with the average spending amount of an individual on new and repeat purchases before the adoption as the baseline. The numbers of observations are different from those in Table 4 because in some observations, the baseline spending amount is zero, making the calculation of proportional changes infeasible.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

A11. Different Pre-Study Windows in Defining New vs. Repeat Purchases**Table A11.1. Moderating Role of New vs Repeat Purchases on Main Effect across All Channels with different pre-study window specifications**

	<i>All Channels (RMB)</i>					
	Pre-study window = 3 Weeks		Pre-study window = 4 Weeks		Pre-study window = 5 Weeks	
	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>
Adoption × Post	0.519 (1.431)	5.200*** (1.428)	0.054 (1.427)	5.665*** (1.432)	-0.017 (1.370)	5.736*** (1.485)
Constant	11.46*** (0.308)	21.47* (0.308)	11.45*** (0.307)	21.48*** (0.308)	11.85*** (0.295)	21.07*** (0.320)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.027	0.127	0.029	0.125	0.034	0.118
	Pre-study window = 6 Weeks		Pre-study window = 7 Weeks		Pre-study window = 8 Weeks	
	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>
	Adoption × Post	0.119 (1.364)	5.600*** (1.492)	0.084 (1.374)	5.635*** (1.484)	0.483 (1.385)
Constant	12.24*** (0.294)	20.68*** (0.321)	12.61*** (0.296)	20.32*** (0.320)	12.99*** (0.298)	19.93*** (0.318)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.036	0.118	0.036	0.120	0.040	0.116

Note: Robust standard errors are in parentheses.
+p<0.1; *p<0.05; **p<0.01; ***p<0.001

Table A11.2. Heterogeneity in the Treatment Effect across Channels with different pre-study window specifications

	Mobile (RMB)		PC (RMB)		Mobile (RMB)		PC (RMB)	
	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>	<i>New Purchase</i>	<i>Repeat Purchase</i>
	Pre-study window = 3 Weeks				Pre-study window = 4 Weeks			
Adoption × Post × Short-term	-0.231 (2.725)	-1.783 (2.734)	0.498 (1.153)	1.170 (1.116)	-0.887 (2.730)	-1.128 (2.726)	0.710 (1.143)	0.959 (1.126)
Adoption × Post × Medium-term	1.279 (1.771)	0.970 (1.776)	0.358 (0.749)	2.475*** (0.725)	0.624 (1.774)	1.625 (1.771)	0.570 (0.743)	2.264** (0.731)
Adoption × Post × Long-term	-0.661 (1.509)	1.146 (1.514)	0.217 (0.639)	1.632** (0.618)	-1.316 (1.512)	1.801 (1.510)	0.428 (0.633)	1.421* (0.623)
Constant	9.070*** (0.281)	18.87*** (0.282)	2.348*** (0.119)	2.555*** (0.115)	9.105*** (0.282)	18.84*** (0.281)	2.299*** (0.118)	2.604*** (0.116)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.023	0.121	0.013	0.065	0.025	0.119	0.009	0.068
	Pre-study window = 5 Weeks				Pre-study window = 6 Weeks			
Adoption × Post × Short-term	-0.100 (2.597)	-1.015 (2.848)	0.776 (1.145)	0.892 (1.125)	-0.906 (2.586)	-1.108 (2.859)	0.819 (1.148)	0.850 (1.124)
Adoption × Post × Medium-term	0.510 (1.687)	1.738 (1.851)	0.636 (0.744)	2.197*** (0.731)	0.604 (1.680)	1.645 (1.857)	0.679 (0.746)	2.155** (0.730)
Adoption × Post × Long-term	-1.430 (1.438)	1.914 (1.577)	0.495 (0.634)	1.354* (0.623)	-1.336 (1.432)	1.821 (1.583)	0.537 (0.636)	1.312* (0.622)
Constant	9.478*** (0.268)	18.47*** (0.294)	2.328*** (0.118)	2.574*** (0.116)	9.810*** (0.267)	18.13*** (0.295)	2.381*** (0.119)	2.521*** (0.116)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.031	0.112	0.009	0.066	0.032	0.114	0.010	0.062

	Pre-study window = 7 Weeks				Pre-study window = 8 Weeks			
Adoption × Post × Short-term	-1.040 (2.587)	-0.975 (2.856)	0.919 (1.171)	0.749 (1.102)	-0.700 (2.599)	-1.314 (2.854)	0.991 (1.204)	0.677 (1.065)
Adoption × Post × Medium-term	0.471 (1.681)	1.778 (1.856)	0.779 (0.761)	2.054** (0.716)	0.810 (1.689)	1.439 (1.854)	0.851 (0.782)	1.982** (0.692)
Adoption × Post × Long-term	-1.469 (1.433)	1.954 (1.582)	0.638 (0.649)	1.211* (0.610)	-1.130 (1.439)	1.615 (1.580)	0.710 (0.667)	1.139 (0.590)
Constant	10.10*** (0.267)	17.84*** (0.295)	2.452*** (0.121)	2.450*** (0.114)	10.43*** (0.268)	17.52*** (0.295)	2.516*** (0.124)	2.387*** (0.110)
Individual FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435	120,435	120,435	120,435	120,435
Adjusted R ²	0.032	0.116	0.012	0.059	0.035	0.110	0.015	0.061

Note: Robust standard errors are in parentheses.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

A12. Channel Dynamics at the Category Level

Table A12.1. Channel Dynamics at the Category Level

	PC Channel (RMB)		Mobile Channel (RMB)	
	<i>Spending on Mobile Refill Cards</i>	<i>Spending on Movie & Sports Tickets</i>	<i>Spending on Online Game Point Cards</i>	<i>Spending on Snacks Nuts & Specialties</i>
Adoption \times Post	-0.258* (0.114)	0.840** (0.278)	-1.438+ (0.748)	0.016+ (0.008)
Constant	0.629*** (0.012)	1.504*** (0.029)	0.920*** (0.178)	0.002* (0.001)
Individual FEs	Yes	Yes	Yes	Yes
Week FEs	Yes	Yes	Yes	Yes
Observations	120,435	120,435	120,435	120,435
Adjusted R ²	0.24	0.11	0.09	0.03

Note: Robust standard errors are in parentheses.

+p<0.1; *p<0.05; **p<0.01; ***p<0.001

We use the “*Movie & Sports Tickets*” category as a representative example to highlight the potential challenges in relying solely on the voice channel for transactions. Given that tasks like seat selection (Step 1) and customer service interactions (Step 2) are involved, it is often more feasible for consumers to transition to the PC channel for finalizing their purchases. This example can help explain the positive spillover effect onto the PC channel.

Table A12.2. Booking Procedure of Movie Tickets

Booking Procedure originally in Chinese (left), translated into English (right)	
<p style="text-align: center; color: red; font-size: 1.2em; font-weight: bold;">订票流程</p> <ol style="list-style-type: none"> 1. 去任意一个卖电影票的app选择你想看的电影场次座位然后截图。 2. 把截图发给客服，客服查询价格后报价给您。 3. 确定后下单。 4. 出票后发你取票码，保存到手机到取票机取票即可。 <p>A real example from Taobao: https://www.taobao.com/list/item/651660707413.htm?spm=a21wu.10013406.taglist-content.1.111a35b9nFHFhl</p>	<p style="text-align: center; font-weight: bold;">Booking Procedure</p> <ol style="list-style-type: none"> 1. Go to any app that sells movie tickets, select the seats for the movie you want to watch and take a screenshot. 2. Send the screenshot to customer service, who will check the price and give you a quote. 3. Place the order after confirmation. 4. After the ticket is issued, you will be sent a ticket collection code, save it to your mobile phone and go to the ticket collection machine to collect the ticket.

A13. An Additional Dataset and the Instrumental Variable Method

The goal of this study is to strengthen the robustness of our main finding that a consumer's Genie adoption ($Adoption_i$) significantly increases the consumer's Tmall spending. Specifically, we supplement the DiD estimation approach in the main text with an instrumental variable (IV) approach.

Data and Background

We obtain an IV from a natural experiment involving a limited-time offer for 60% off the Genie device. The coupon was displayed prominently on the Genie shopping page from January 25 to February 20, 2019, such that every consumer who went to the Genie online store during this period presumably saw the coupon. The platform did not inform customers in advance about the coupon, so no customers should have had information about the upcoming discount prior to January 25, 2019. We use exposure to the coupon as an intention to treat (ITT_i) since the coupon encouraged consumers to purchase Genie and thus should correlate with Genie adoption.

To construct the dataset, we randomly draw from the pool of customers who visited the Tmall Genie official store during either January 23–24 (before the coupon was available) or January 25–26 (when the coupon was first available).⁴⁴ The dataset includes each consumer's demographics as well as all Tmall transactions from November 30, 2018, to March 7, 2019 (13 weeks). As in the main text, we aggregate the data at the customer-week level, yielding 8 weeks of pre-treatment observations (from November 30, 2018, to January 22, 2019) and 5 weeks of post-treatment observations (from January 27 to March 7, 2019). We are interested in Genie adoption decisions that occurred from January 23 to January 26, 2019. We exclude consumers who did not visit the store between January 23 and January 26 or who visited during this period but purchased

⁴⁴ Here, a "customer" refers to a general Alibaba account holder.

the Genie before January 23 or after January 26. We also exclude customers who visited the store on both January 23–24 and January 25–26 because these customers are rare and seem unusually active, raising a potential selection issue.

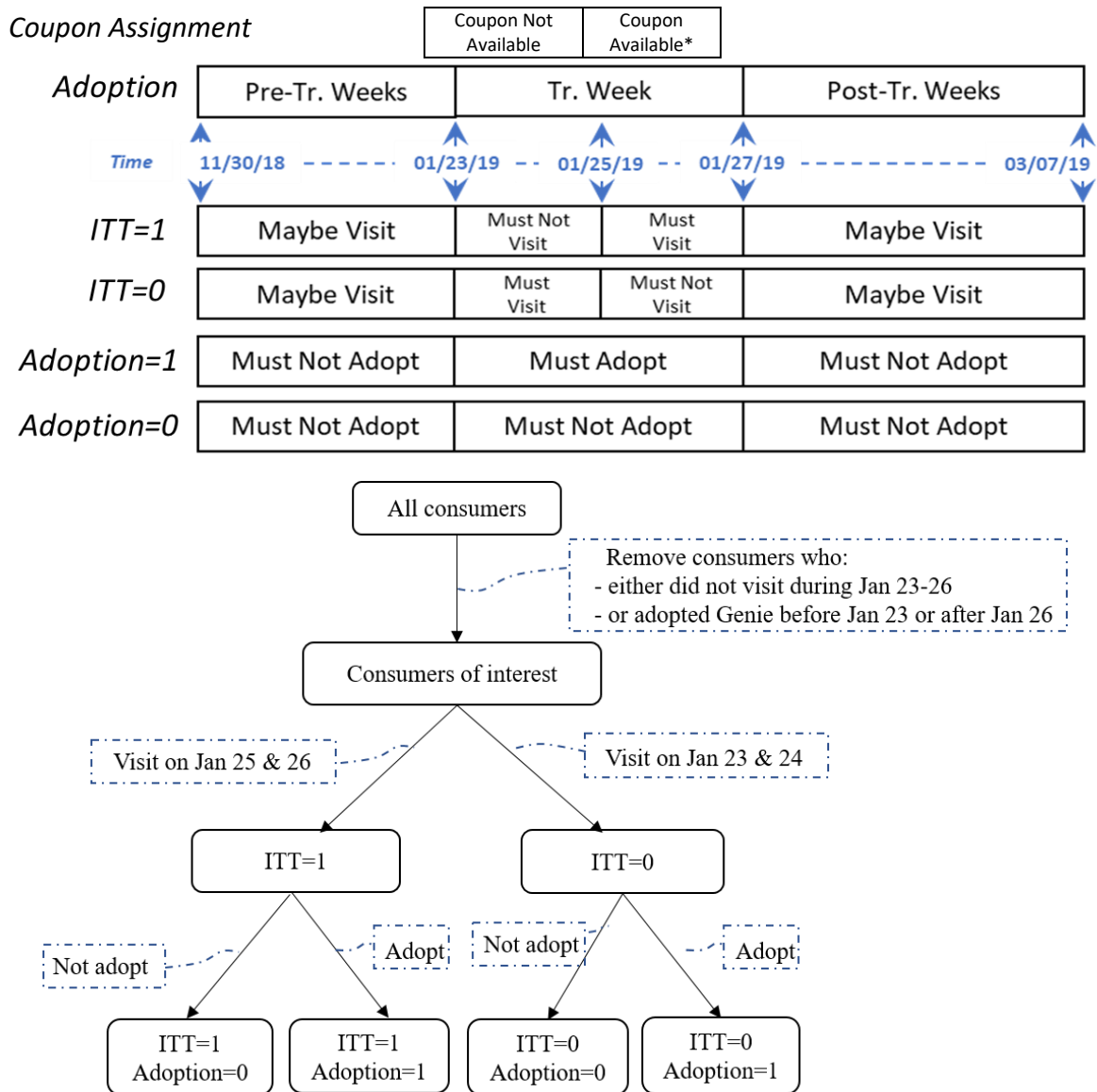
We assign $ITT_i = 1$ to customers who visited the store on January 25–26 (and presumably were exposed to the coupon), and we assign $ITT_i = 0$ to customers who visited on January 23–24 (and could not have seen or used the coupon). The assignment of ITT_i is thus random because the coupon launch date was determined exogenously by the platform, as we confirmed with the platform that no customers were informed in advance about the release of the coupon. Consumers who visited in the 48 hours before 12:00 AM on January 25 should not be systematically different from those who visited in the 48 hours after. In other words, coupon exposure is uncorrelated with unobserved consumer characteristics, yet consumers with coupon exposure (relative to those without exposure) should be more likely to adopt the Genie. Thus, we expect coupon exposure to meet the requirements for a valid IV (explained in Analysis and Results, below).

We pool consumers into four types, visualized in Figure A13. Type 1 is a coupon-receiver and Genie adopter (for whom $ITT = 1$ and $Adoption = 1$). Type 2 is a non-coupon-receiver and non-adopter (for whom $ITT = 0$ and $Adoption = 0$). Type 3 is a coupon-receiver and non-adopter (for whom $ITT = 1$ and $Adoption = 0$). Type 4 is a non-coupon-receiver and adopter (for whom $ITT = 0$ and $Adoption = 1$).

Due to the sample size restriction imposed by the company, we down-sample the original dataset for subsequent analysis. For each group, we down-sample with respect to ITT —that is, our down-sampling occurs at the ITT level but not within it, because ITT is randomized while $Adoption$ given ITT reflects the compliance rate, which is not random. Specifically, we down-sample ($ITT = 1, Adoption = 1$) from 8,312 to 1,000; ($ITT = 1, Adoption = 0$) from 290,920 to 35,000; ($ITT = 0,$

$Adoption = 1$) from 19,850 to 200; and ($ITT = 0, Adoption = 0$) from 4,400,745 to 44,340.⁴⁵ We use a cloud-based tool provided by Alibaba to ensure that the customer sampling process is purely random and that the selected consumers are representative of the overall population. Table A13.1 summarizes some key variables for the resulting sample.

Figure A13. Visualization of the Empirical Setting



⁴⁵ We start with ($ITT = 1, Adoption = 1$), which has 8,312 consumers, and we down-sample to 1,000. We apply the same down-sample ratio (8,312) to the other $ITT = 1$ group: 290,920 consumers / 8,312 = 35,000 consumers in the down-sample of ($ITT = 1, Adoption = 0$). We follow a similar procedure to down-sample the groups for which $ITT = 0$.

	<i>Adoption=1</i>	<i>Adoption=0</i>
<i>ITT=0</i>	Coupon-receiver and Purchaser (1,000 customers)	Coupon-receiver but Non-purchaser (35,000 customers)
<i>ITT=1</i>	Non-coupon-receiver but Purchaser (200 customers)	Non-coupon-receiver and Non-purchaser (44,340 customers)

Note: *The box reports that coupons are available until only January 26 because we consider only the consumers who visited between January 23, 2019 and January 26, 2019 (even though coupons actually were available until February 20).

Table A13.1. Summary Statistics

Feature/Variable	Mean	S.D.	Min.	Max.
<i>Demographics</i>				
Age	29.46	8.239	18	87
Gender	0.49	0.50	0	1
<i>Dependent Variables</i>				
Total Spending Amount (RMB)	311.4	1,677	0	300,585

Analyses and Results

Recall that our goal is to estimate the ATT of a consumer's Genie adoption on the consumer's small spending. Consistent with our analysis in the main text, our primary outcome variable is the *spending amount* (y_{it} , the amount spent by consumer i in week t). Note that coupon exposure itself is an intention to treat (*ITT*), and it is correlated with the treatment (*Adoption*) but not equivalent to treatment ($ITT_i \neq Adoption_i$) due to the two-sided noncompliance issue (Imbens and Rubin, 2015). That is, consumers who were assigned to the intention-to-treat group (i.e., consumers who visited on January 25–26 and thus presumably saw the coupon) may have chosen not to adopt the Genie; also, consumers who were not in the intention-to-treat group may have chosen to adopt the Genie even though they did not have a coupon. To address the two-sided noncompliance issue, we leverage the assignment of *ITT* (i.e., coupon exposure) as the IV for treatment (Chapter 24 of Imbens and Rubin 2015).

A valid IV must satisfy three assumptions. In the *relevance* assumption, the IV significantly affects the independent variable. In our case, ITT_i significantly affects $Adoption_i$, and our IV passes the weak instruments test (F-statistics = 597.679). In the *monotonicity* assumption, seeing the coupon must not decrease the consumer's spending. In our case, the coupon itself (whether redeemed or not) did not contain any information that should have discouraged consumers from purchasing or searching for other products.

In the *exclusion restriction* assumption, ITT_i should affect the outcome variable only via $Adoption_i$. Specifically, the IV has to be exogenous (in our case, coupon access is randomly assigned and uncorrelated with the other variables), and the IV must not directly affect the outcome variable. In our case, seeing the coupon without using it to adopt the Genie must not affect the consumer's subsequent spending. We believe this assumption is satisfied for three reasons. First, the coupon provides a discount only for the Genie, not for other products, so the coupon should not discourage or encourage consumers to purchase other products. Second, without redemption (i.e., without using the coupon to purchase the Genie at the discounted price), the coupon itself is of no use to the consumer, so it cannot directly affect y_{it} . Third, coupon data are not shared among sellers on the platform, so they cannot target consumers who were exposed to the coupon.

To further make sure there is no significant difference between the control group and the treated group, we first remove 1% outliers (the top 1% in weekly spending) and apply propensity score matching (PSM) before adopting the IV approach. Specifically, we follow the same procedure of PSM as in the main text, matching on users' demographics (gender and age) and pre-treatment purchase behavior (i.e., the number of previous purchases and spending). As shown in Table A13.2 and Figure A13.2, after matching, the treatment and control groups are similar in their

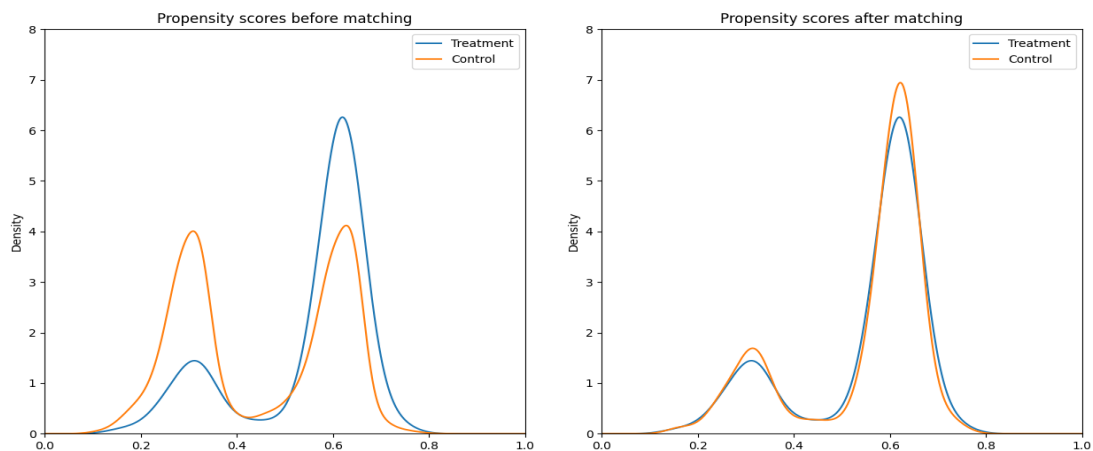
adoption propensities (Figure A13.2), and there is no significant difference between the two groups in demographics and pre-treatment shopping behaviors.

Table A13.2. Comparison of Adopters and Non-Adopters Before and After Matching

	Adopter (N)	Adopter (Mean)	Non-adopter (N)	Non-adopter (Mean)	p-value of t-test
(a) Before matching					
Age	1,144	27.738	71,135	29.426	0.000
Gender (Male = 1)	1,144	1.795	71,135	1.513	0.000
Number of Purchase items	1,144	280.589	71,135	310.846	0.019
Spending_Amount	1,144	252.349	71,135	241.069	0.270
(b) After matching					
Age	1,144	27.738	5,402	27.720	0.935
Gender (Male = 1)	1,144	1.795	5,402	1.784	0.419
Number of Purchase items	1,144	280.589	5,402	280.657	0.812
Spending_Amount	1,144	252.349	5,402	244.040	0.474

Note: matching was conducted after 1% outliers being removed.

Figure A13.2. Distribution of Propensity Scores Before and After Matching



Next, we describe the estimation of the ATT of Genie adoption on spending. IV estimates usually are calculated with the two-stage least squares approach (2SLS). Typically, in the first stage, the endogenous variable (*Adoption*) is regressed on the instrument and all exogenous covariates in the equation of interest, and the predicted values are obtained. In the second stage, the regression of interest is estimated as usual, but each endogenous covariate is replaced with the predicted value. In our analysis, the endogenous variable is categorical, so we cannot use ordinary

least squares (OLS) for the first stage; instead, we estimate a Probit model using the Chamberlain-Mundlak device. In the second stage, we regress Y on the predicted values from the first stage with a linear model (OLS estimates).⁴⁶ Given the extremely large maximum (compared to the mean) weekly spending amount (see Table A13.1), we removed outliers before the estimation. Table A13.3 shows the results after removing outliers in the top 1% and top 5% of weekly spending.

We find that Genie adoption leads consumers to spend an average of RMB 266.3 more every week.⁴⁷ The significant, positive treatment effect is qualitatively consistent with the DiD estimation results. The magnitude of the effect is larger here, which we suspect is attributable to seasonal effects, the sample period includes the Lunar New Year, when consumers spend much more than usual.

Table A13.3. ATT of Genie Adoption on Spending

	<i>Spending (RMB)</i>	
	1% outliers removed	5% outliers removed
Treatment Effect	226.3*** (24.43)	182.5*** (25.13)
Controls	Y	Y
Individual FEs	Y	Y
Week FEs	Y	Y
Observations	85,098	77,857

Note: z-statistics are in parentheses; *** < 0.001 ** < 0.01 * < 0.05

We acknowledge that a consumer who purchases the Genie should have a smaller budget for other consumption on the platform. Thus, the estimated positive effect of Genie adoption on spending would be larger without a budget effect (i.e., the budget effect is absorbed in the treatment effect). As we do not observe whether each consumer holds a fixed or flexible budget for their consumption on Tmall, we cannot precisely disentangle the budget effect from the treatment effect.

⁴⁶ <https://www.statalist.org/forums/forum/general-stata-discussion/general/1356744-binary-endogenous-treatment-in-panel-data>.

⁴⁷ We converted the spending amount from RMB to US dollars using an exchange rate of 7.