Title
Language translation effects in Chatbots: Evidence from a randomized field experiment on a mobile commerce platform

Authors
Nayak, Ashutosh
Nair, Ashwin Aravindakshan

Publication Date
2025-03-01

DOI
10.1016/j.jbusres.2024.115158

Highlights

Language Translation Effects in Chatbots: Evidence from a Randomized Field Experiment on a Mobile Commerce Platform

Ashutosh Nayak, Ashwin Aravindakshan Nair

- Research Focus: Investigates the impact of language localization in AI chatbots on user behavior in a mobile commerce platform in India, particularly among bilingual users.
- Research Gap: Addresses the lack of research on the mixed effects of language localization in markets where English is an official but not the sole language.
- Theoretical Framework: Utilizes the Technology Acceptance Model (TAM), Unified Theory of Acceptance and Use of Technology (UTAUT), and Cognitive Load Theory (CLT) to explain user interactions with language-switching chatbots.
- Methodology: Conducted a six-day randomized field experiment where users were assigned to either an English-only version or a bilingual (English-Hindi) version of the app to evaluate changes in purchases and uninstalls.

Language Translation Effects in Chatbots: Evidence from a Randomized Field Experiment on a Mobile Commerce Platform

Ashutosh Nayak[a], Ashwin Aravindakshan Nair[b]

[a] *Data Intelligence Lab, Samsung Research Institute Bangalore, Bengaluru, 560066, Karnataka, India*
[b] *Department of Managament Studies, Uniersity of California, Davis, Davis, 95616, California, USA*

Abstract

This study investigates the impact of language translation innovations in artificial intelligence (AI) digital assistants. We use data from a mobile commerce platform application in India that introduced a Hindi and English version of its previously English-only language chatbot. The data, obtained from a randomized field experiment conducted on new users, help determine the impact of introducing language translation in conversational chatbots by quantifying its effect on user metrics such as purchases and uninstalls. In the experiment, the firm only altered the language of interaction, from English-only to Hindi and English. The firm did not make any other changes in the design or purchase flow within the application. We find that language translation innovations significantly increase the number of user sessions and also improve user purchases and engagement. The increase in the engagement did not emerge from an increase in the number of sessions but from an increase in interactions within a session in the bilingual app. We also observe a sharp rise in uninstalls for the population that received the bilingual app. We find that in the bilingual chatbot, uninstalls rise with increased user interactions in a high involvement product category. In sum, the results from the field experiment show that while language translation in artificially intelligent assistants leads to greater purchases, it could also lead to increased uninstalls. This result suggests that implementing similar language translation innovations in isolation, without any modifications to in-application experience, has the potential to yield negative outcomes for the firm.

Keywords: Artificial intelligence, Conversational Chatbot, Mobile apps, Localization, Online shopping behavior

1. Introduction

Recent advances in natural language processing have led to the increased use of chatbots or AI assistants in electronic and mobile commerce. Chatbot-enabled commerce is expected to exceed $14 billion by 2024 [1]. [2] argue that the "first domains to be affected by AI are likely to be settings in which AI systems can be seamlessly embedded into existing systems." Chatbots exemplify this argument. Chatbot use encompasses several facets of the customer-firm relationship, from customer service to facilitating purchases (see Figure 1).

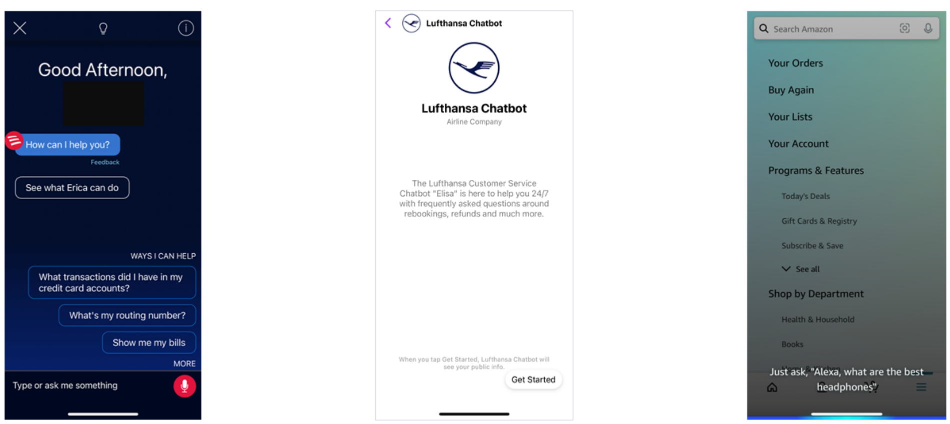


Figure 1: Examples of Chatbots Embedded in Mobile Applications for Bank of America, Lufthansa and Amazon

As firms globalize and expand chatbot usage into new markets, they innovate on multiple aspects of chatbot design and interaction, including language. In most countries, chatbots have traditionally conversed in English. However, as firms move into markets where English is not the primary language, localization becomes essential. In countries like India, where English is one of the official languages but not the sole language of communication, it is crucial to consider the nuances of bilingual or multilingual interaction. While English is commonly used in formal and business settings, a significant portion of the population prefers communicating in regional languages, especially in personal and informal transactions. For example, the majority of new internet users in India are non-English speakers, with estimates indicating that 57% of internet users prefer to browse in their local Indic language [1]. This underscores the importance of studying the effects of language localization, even in multilingual markets like India, where supporting both Hindi and English in the same conversation may offer a more seamless experience for users. Apps invest in localization on the promise of downstream economic benefits [3], but how these benefits manifest at the user level, especially in a bilingual context, remains underexplored.

This study addresses the gap by exploring how user behavior changes when language localization through AI agents, such as chatbots, is incorporated into mobile commerce platforms. While localization is generally seen as beneficial, leading to an increase in purchases ([4], [5], [6], [7], [8], [9]), could there also exist negative outcomes associated with localization efforts that focus only on language? Indeed, the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) further suggest that language translation can positively impact business outcomes. These models argue that the perceived ease of use and usefulness of a system increase when users can interact in their preferred language, leading to higher adoption rates and user engagement. Recent work by [10] demonstrated that introducing machine translation on e-commerce platforms like eBay significantly boosted international trade, reinforcing the potential benefits of language localization.

However, we also explore whether language localization could yield adverse outcomes, as suggested by some studies on bilingual advertisements, where shifts in language can affect user trust ([11], [12]). Furthermore, Cognitive Load Theory (CLT) introduces a potential tension by suggesting that while language translation may enhance user experience in some instances, it could also increase the complexity of interactions. For high-involvement

[1]`https://uat.indiadigitalsummit.in/sites/default/files/thought-leadership/pdf/Kantar_iamai_Report_20_Page_V3_FINAL_web_0.pdf` (Accessed on 09-26-2024.)

transactions, switching between languages or navigating complex tasks in a vernacular language might increase cognitive load, leading to frustration and negative outcomes. Additionally, the increased anthropomorphism of a chatbot through vernacular use may raise users' expectations of its competence, potentially leading to frustration if these expectations are not met ([13]). Therefore, our study fills these gaps by examining these nuanced effects of language localization, especially in the context of India's bilingual users who may face differing experiences when interacting with a language-switching chatbot.

In this context, we seek to answer the following research questions:

- How does the introduction of language localization in a chatbot impact user purchases?
- How does the introduction of language localization in a chatbot impact user retention?
- Does the complexity of purchase transactions affect how users respond to language localization when interacting with a chatbot?

We answer these questions through a six-day randomized field experiment conducted with a mobile commerce platform in India. Although English is widely spoken, many users prefer interacting in Hindi or switching between languages depending on context. The platform offers various products, classified into low-involvement (utility payments) and high-involvement (travel bookings, local deals) categories. New users were randomly assigned to download either Version 1 (V1, English-only) or Version 2 (V2, Hindi and English) of the app. The V1 chatbot conversed only in English, while the V2 chatbot could switch between Hindi and English within the same conversation. Users were unaware of the new language feature prior to installation, ensuring unbiased interaction.

To evaluate the effects of language localization, we tracked key performance metrics, including purchases, uninstalls, and app interaction (text exchanges with the chatbot). Model-free evidence shows that purchases and interactions per session increased by 146% and 132%, respectively, in V2 compared to V1. However, the uninstall rate for V2 was 107% higher than V1. Further analysis, accounting for factors like transaction complexity, attributes 87% of the purchase increase and 76% of the uninstall increase to the impact of localization. Specifically, we find that low-involvement transactions benefited from localization, while high-involvement transactions, which require more user input and longer paths to purchase, led to higher user churn. This suggests that language localization, if implemented without adjustments to the overall user experience, can negatively impact user retention for complex transactions.

These results offer a nuanced view of language localization in chatbot interactions. While it can enhance user engagement and purchases, firms must carefully design localized experiences to avoid potential pitfalls such as increased cognitive load and user frustration. This study builds on existing literature by empirically quantifying the effects of a bilingual AI assistant on multiple user metrics, providing insights that can guide future localization efforts. To the best of our knowledge, this is the first study to examine these effects in a conversational AI context.

The remainder of the paper is organized as follows. In Section 2, we review the current literature and build a theoretical framework for studying app localization effects on chatbots, integrating theories such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT). Section 3 outlines the research setting and experiment. Section 4 presents the data, followed by model development in Section 5. Section 6 discusses empirical results, with extensions and managerial implications in Section 7. Finally, Section 9 concludes the study.

2. Literature Review and Theoretical Development

In this study, we investigate the effect of language innovations in chatbots employed by mobile retail platforms. Recent research on AI chatbots in mobile commerce emphasizes their growing influence on customer experience and business outcomes. For example, [14] underscores the importance of chatbot anthropomorphism and design in influencing consumer behavior, showing that features like gendered personas and human-like traits like empathy ([15]), interactivity, and narrativity ([16]) can significantly shape user trust and engagement. Additionally, [17] highlights the transformative role of AI-powered conversational agents across industries, emphasizing the need for nuanced approaches to design and integration that consider user preferences and behavioral impacts. Building on this, studies exploring language and cultural dynamics in chatbot interactions reveal that localization and context-specific adaptations are critical for enhancing user satisfaction, though they may introduce cognitive challenges in complex tasks. [18] highlight the importance of language variation in chatbot interactions, finding that using register-specific language tailored to social contexts enhances user perceptions of appropriateness, credibility, and overall experience. Similarly, [13] examine how anthropomorphism in chatbots affects customer emotions, revealing that overly human-like chatbots can intensify negative reactions when user expectations are not met. This underscores the need for thoughtful design to manage customer interactions effectively. Complementing these findings, [19] systematically reviews AI conversational agents (CAs), including chatbots, across industries such as retail, banking, tourism, and healthcare, identifying key research areas like consumers' trust, Natural Language Processing (NLP) in chatbot design, communication dynamics, and the impact on business value creation. The study highlights the significance of customization and personalization in chatbot design, suggesting that adapting CAs to user preferences, cultural contexts, and language styles is vital for enhancing trust and engagement. Collectively, these findings underscore the need for language-based adaptations, implying a future research direction focused on multilingual CAs and their impacts on business outcomes.

In this study, we address this gap by determining the contribution of language localization efforts in chatbots towards increasing sales and enhancing customer relationships. Most of the advancements in language localization currently lie in the space of *static translation* of text within the app into the vernacular of specific users. For example, using data from eBay, a recent study by [10] shows that machine translation helps improve commerce in a highly globalized world by connecting shoppers across multiple geographies to websites in their local languages. Unlike machine translation, however, effective language localization in an app using chatbots needs significant resources from the firm not only in terms of the translation technology but also to understand context, culture and communicate with its users in their local languages in real-time. Firms adopt these innovative technologies with the expectation of downstream economic benefits such as increase in sales or retention in the app e.g., market expansion to a non-English speaking populace. We quantify its impact and study how the innovations in chatbot technology affects metrics of user behavior important to the firm – purchases and retention.

Business practice and research have conducted several studies on the benefits of localization. [4] documents the benefits of universal advertising appeals communicated in local languages. [5] discusses the positive effect of the local language on a user's memory. Additionally, [6] studied the persistence of this effect when changing languages within a conversation in marketing messaging. [7] studied the emotional intensity of local languages, translating to improved user perception. More recently, [8] show that native-language advertisements elicit self-referent thoughts about family, friends, or home. This then leads to more positive attitude measures and behavioral intentions. [9] also showed the positive sales effects of salespersons conversing in a regional dialect.

In addition to this, the introduction of a vernacular language within the app could increase the perceived usefulness (PU) and perceived ease of use (PEOU) of the app. These benefits, as described by the Technology Acceptance Model (TAM) are pivotal in influencing an individual's attitude toward using a technology, which subsequently shapes their behavioral intention and actual usage ([20], [21]). In addition, the vernacular version could also foster an emotional attachment to the bot that in turn leads to social acceptance of the technology. Thus, following the Unified Theory of Acceptance and Use of Technology (UTAUT) proposed by [22], one would also expect a positive link between the introduction of the vernacular version and the purchase outcomes. Both the TAM and UTAUT theories suggest that perceived ease of use and usefulness of a system, such as a language-switching chatbot, lead to higher adoption and engagement rates. In sum, based on the existing research, and the expected positive impact of language localization, we hypothesize that:

Hypothesis 1. Introducing a vernacular version of the app with capability to converse in the local language will increase the average number of consumer purchases within the app.

Users' response to localization could also depend on the nature of the transaction. Cognitive load theory (CLT) ([23], [24]) posits that the human cognitive system has limited capacity, and when overloaded, learning and performance suffer. For example, [25] find that, in simpler less complicated transaction settings, online automated agents enhance the user experience and positively influence the firm outcomes. Similarly, adding a new language introduces a new mode of interaction with the chatbot. This new interaction mode may enhance the level of experience in the mobile app. Recent studies in anthropomorphism find that firms should start with the automation of easier tasks when introducing anthropomorphic agents, at least until the innovative technology is matured. For example, [26] suggest that during the exploration stage of the new technology, a positive confirmation regarding an avatar's behavioral realism can ensure good cognitive experiences. [27] posit that conversational AI's user engagement hinges on its human-like competencies: cognitive, relational, and emotional. Cognitive competency enhances user reliance by reducing cognitive effort, while relational competency fosters cooperation and lessens communication ambiguities. Emotional competency connects with users on an emotional level, offering warmth and empathy. Crucially, these competencies increase user trust in AI, which mediates the relationship between AI competencies and user engagement, making trust a pivotal factor in effective AI-user interactions. In addition, [28] reveal that chatbots with advanced conversational skills, demonstrated by tailored responses and response variety, are perceived as more socially present and human-like. This enhanced social presence directly leads to higher user engagement and perceived humanness. The study emphasizes the significant role of conversational abilities in influencing user interactions, highlighting how skilled chatbots foster both mindless and mindful anthropomorphism among users. Thus, firms should start with easier tasks which the nascent technology such as language translation can complete with less errors, enhancing trust in the technology. We expect that this will lead to a positive user experience with the vernacular AI agents, and in turn an increase engagement, the probability of purchase, and decrease the probability of uninstall. For more complex transactions or experiences, the higher magnitude of anthropomorphism may be detrimental to the consumer's in-app experience due to the increased cognitive load on the consumer. Additionally, CLT raises the possibility that adding language options in high-involvement transactions may increase user frustration. This in turn could have a negative effect on the consumers' purchase decisions and their retention.

Language localization efforts without any meaningful changes in other aspects of the purchase process can also interact with the complexity of the transaction. For example, in the presence of detailed information, the agents could negatively impact the user experience. [25] find that AI agents need to first gain users' confidence through discrete and basic exchanges, before user can expect a richer and more involved online transactions in complex service settings. Combining the effect of potential negative effect of language localization, unexpected changes in user experience and interaction of complex product purchase with language translation in AI assistants, we expect that not all users will respond favorably to the introduction of new language in the app. Thus, we hypothesize that:

Hypothesis 2A. Language localization in chatbots will yield more positive outcomes for less involved and less complex product transactions.

Hypothesis 2B. Language localization could lead to less desirable results when customers deal with highly involved and complex product transactions.

Finally, past literature also shows that under certain circumstances language localization may not yield positive outcomes. For example, [11] found that advertising exclusively in Spanish to a target Hispanic population decreased affect towards the advertisement. The reason that this occurs because the exclusive use of Spanish in advertising could arouse Hispanic insecurities about language usage. Similarly, in the Indian context, [12] study the role of advertising languages in countries where the population is bilingual. They focus their study on the urban Indian population – proficient in both English and Hindi – a population that also forms the majority of app users in India. They find that the *unexpectedness* of Hindi language choice focuses more attention on the language of the ad rather than the ad itself. This then heightens the viewer's skepticism, leading to increased counterarguments, in turn, reducing the ad's effectiveness. These studies appear to suggest that firms should observe caution when localizing their chatbots. For example, would users familiar with conversing in English in the app suspect a bilingual update? Would users unaware of the bilingual bot find the unexpectedness of the conversation detrimental to their in-app experience? Localization, in many cases, could help expand the market, but it could also potentially turn off new and existing users who find the use of the language to be unexpected in the context of AI agents. We expect language localization to impact purchase outcomes positively. However, the possible unexpectedness of or increased involvement in the new language could lead to greater user aggravation with the chatbot and possibly worse outcomes like increased customer churn for the firm. For example, [13] shows across five studies and real-world experiments that human-like traits in chatbots (such as names or avatars) lead to lower satisfaction, poorer firm evaluations, and reduced purchase intentions. This is attributed to expectancy violations—customers expect human-like chatbots to be more competent, and when these expectations are not met, their frustration intensifies. This leads us to our next hypothesis:

Hypothesis 3. The unexpectedness of language localization could lead to worse customer experiences which could in turn affect customer metrics like retention.

In sum, we hypothesize that language localization, by itself, could have both advantages and disadvantages for the implementing firm. The findings of this study could help the managers in understanding the impact of localization on their firm before investing in the AI technology. The conceptual framework for this study is shown in Figure 2. We connect two consumer decisions – the probability of purchasing within the app and the probability of uninstalling the app to the instance of whether the consumers used the V2 version of the app with the local language or not. In other words, we measure how the introduction of a chatbot that speaks the local language affects consumer purchase or uninstall decisions. In addition to the mediating role of chatbot language, the difficulty of buying a product also plays a mediating role on total purchases and uninstalls. Specifically, we investigate whether the language of purchase could be affected by the difficulty of the purchase process. Apart from these two mediators, the characteristics of the consumer, mobile phone device and the time of the day are also assumed to play a significant moderating role in determining the behavior of the consumer with the introduction of new language in the app.

In the following section, we provide more details on the research setting and discuss the localization and experiment setup.

3. Research Setting

3.1. The Mobile Platform

We conducted this study in collaboration with a mobile commerce platform in India in September 2020. The platform sells its products exclusively using a mobile app based chatbot. The app uses the state-of-the-art Google API to both understand textual inputs from the user and respond to their queries or answers in English or their local language. Note that the users can make purchases only by interacting with the mobile app as shown in Figure 3. The firm operated the English-version of the app in India for a period of almost two years prior to the experiment. At the time of the study, it had an installed user base of about three million users.

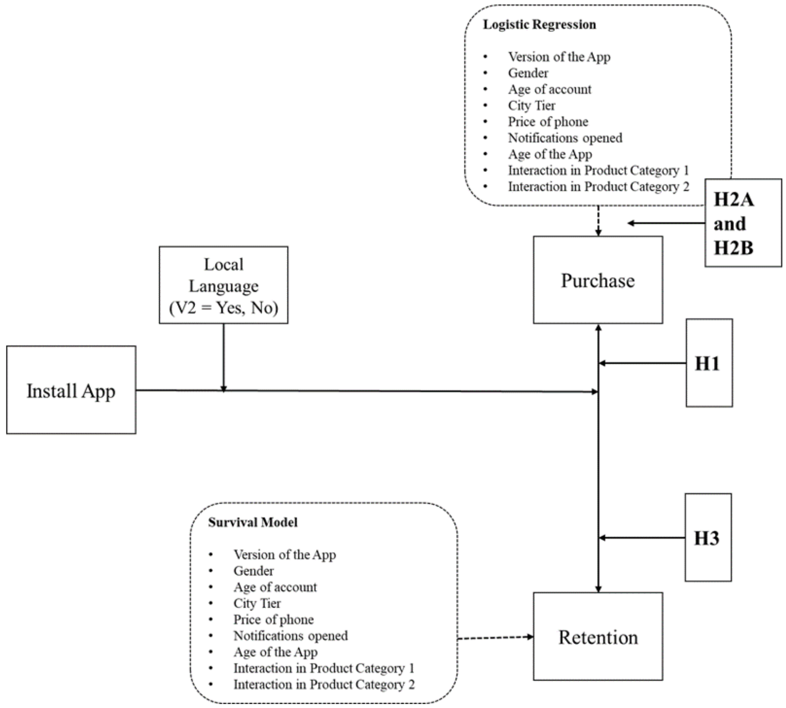


Figure 2: Conceptual model to identify the impact of language localization in mobile app

The app platform is powered by an artificial intelligence-driven chatbot that functions as an intelligent personal assistant. Using natural language processing and machine learning, it offers a chat-based interface, allowing users to interact in their preferred language. The chatbot is designed to understand conversational nuances specific to India, interpret users' requests in the context of products or services, and provide tailored recommendations accordingly.

The app platform offers multiple products. We classify these products into two categories: (1) a low involvement product category, $Category_1$, that includes utility bills payment (electricity, cable/direct-to-home, water, gas) and mobile phone plans (prepaid and post-paid) and (2) a high involvement product category, $Category_2$, that includes travel booking (car, bus, movies, events, and hotels) and other products (local deals, gift cards, and grocery). $Category_1$ includes services with a straightforward path to a transaction in the app – i.e., search for the service and complete payment by selecting a payment gateway. These include actions that are mostly automated and need minimal or limited input and interaction from the user. $Category_2$ includes products that require more instructions for the chatbot to complete the transaction, about 64 characters versus 37 characters for $Category_1$. These instructions could vary depending on the product and may require more search, for example, looking for deals, or browsing product options. $Category_2$ generally has more options to select from and requires the user to interact in a more involved manner with the chatbot. $Category_1$ products have a more straightforward purchase process, mainly involving bill payments.

3.2. Localization and Indian Context

The Indian government recognizes English and Hindi along with several other regional languages as official government languages at the state or federal level. Many families in the middle and upper strata of Indian society are bilingual, conversing in Hindi (or the regional language) and English in social settings. In general, the English-speaking population includes urban (and suburban) areas and primarily the educated and more advantaged strata of society ([29]). This population also tends to be early adopters of technology products – hence, the primary initial target of the app with which we collaborated. However, while English is an official language in India, it is only spoken by a little over 10% of the population [2].

India's vast linguistic diversity offers a strong rationale for developing a local language app. With over 1,600 languages spoken and 22 officially recognized languages, India's population communicates in a wide variety of languages, each tied to its regional identity and culture. While Hindi is the most widely spoken language, a significant portion of the population prefers conducting daily transactions in their native tongues, especially in rural and semi-urban areas. English, though common in business and education, does not resonate with all users, particularly those less familiar with it. By offering a local language app, companies can tap into a broader audience, enhance user engagement, and provide a more personalized experience. Additionally, due to a growing demand for smart devices and the decreasing cost of smartphones and mobile data, firms can no longer afford to ignore the portion of the market that does not converse in English. To satisfy this demand and to

[2] `https://censusindia.gov.in/nada/index.php/catalog/42561,https://www.livemint.com/news/india/in-india-who-speaks-in-english-and-where-1557814101428.html`(Accessed on 09-19-2024.)

(a) Part 1: Sample English Chat

(b) Part 2: Sample English Chat

(c) Part 1: Sample Hindi and English Chat

(d) Part 2: Sample Hindi and English Chat

Figure 3: A sample transaction in English (a and b) and Hindi (c and d). Blue bubbles denote the user inputs and white bubbles indicate the chatbot replies.

cater to India's language diversity, the focal firm introduced a new bilingual (English and Hindi) version of the in-app chatbot. Figure 3 displays a sample interaction with the app for mobile phone plan purchase transactions within the mobile app in English and Hindi.

We note that the act of introducing a bilingual bot is different from simply incorporating machine translation technology. The bot must not only understand the language the user converses in but also understand the context with regards to the transaction and engage in a conversation with the user. In some cases, the users also transition between Hindi and English in the same transaction. Apart from the financial costs, changing the user experience in such dramatic ways can negatively impact the user, thus turning away loyal users who previously interacted with the app in English only versions. Given this risk and investment in developing the conversational experience, the firm sought to quantify the economic impact of this localization to determine if the effort delivers a positive return. We accomplish this using a randomized field experiment, which we discuss next.

3.3. Randomized Field Experiment Setup

To study the impact of this localization in the mobile commerce platform, the firm ran an experiment that spanned six days. In the experiment, the firm randomly assigned a new user to version V1 or V2 of the app. Specifically, during the period of the experiment, when new users chose to *install* the app from the app store, they had a two-thirds probability of downloading V1 and a one-third probability of downloading V2. If a user got V2, they only learned about the bilingual features after installing and opening the app. Thus, a user would not know that the app supports Hindi prior to this step, even if they were assigned to the V2 condition. We classify users into six different cohorts, one for each day of the experiment depending on when they installed the app. Users within a cohort installed the app on the same day of the experiment. Users who installed the app on Day 1 of the experiment were in Cohort 0 and Day 6 in Cohort 5.

A user with V2 of the app could interact with the app in English or Hindi. The app uses Google API for language translation. For the period of the experiment, the firm did not send any personalized marketing campaigns to these users. Additionally, before and during the experiment period, the company did not advertise its new language feature. Figure 4 illustrates the design for the experiment. It also shows the number of users who installed the app on different days during the period of the experiment. In total, over the six days period, 8,683 individuals received V1 (control group) and 3,952 individuals received V2 (treatment group) for a total of 12,635 users who installed the app over the 6 days. After installation, we tracked the behavior of the new users for these six days (or the remainder thereof). On the seventh day, the firm allowed all the users to upgrade to V2 of the app. They sent a notification one day after the experiment to all the users with V1 that a new version is available.

To isolate the effect of the new language introduction, for the period of the experiment, the firm did not make any other design changes to the app. To maintain the sanctity of the experiment and avoid data contamination issues, we consider the data from the first six days of the experiment. The six-day length of the experiment also allows us to avoid cross-contamination caused by word-of-mouth about the bilingual version of the app. For this study, we consider that an user uninstalled the app when they use the uninstall option in the app. However, some users can silently stop using the app, but due to short period of the study, we do not count such users as those who uninstalled the app. Finally, we note that no personal identifying information about the individual is present in the dataset. We also ensure this by checking that none of the 12,635 users considered in the study were referred by the existing users. We describe the data collected in the next section and then proceed to the analysis and results.

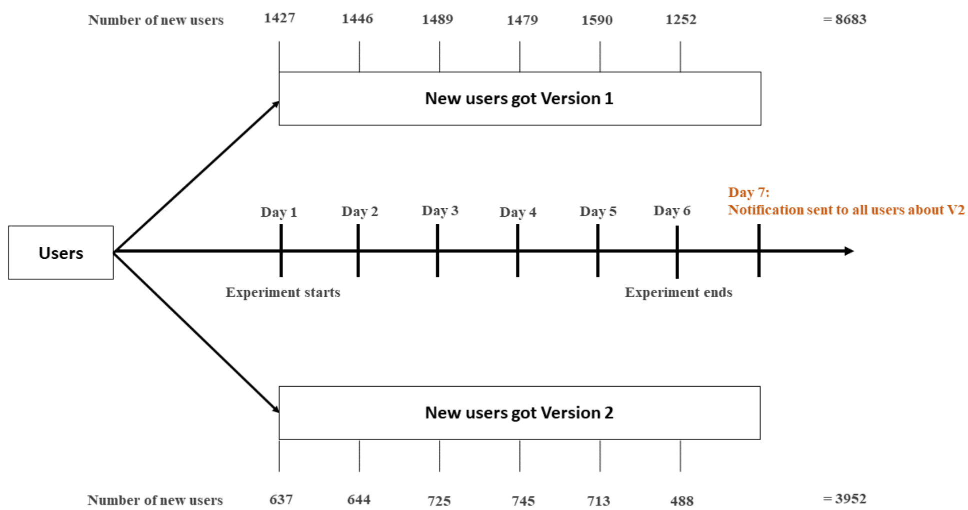


Figure 4: Design of the Randomized Field Experiment

4. Data and Initial Exploration

4.1. Data Description

In this study, we aim to understand the observed impact of language translation in the mobile commerce app. We consider two different outcome variables to quantify this impact. We also consider different factors that could provide a better understanding of the observed outcomes. Tables 1 and 2 provide an overview of the data collected during the experiment. Next, we discuss the two outcome variables and different factors used as control variables in the linear models before presenting the model-free evidence.

4.1.1. Outcome Variables

We measure two metrics at the level of the user: (1) Daily Purchases; and (2) Uninstalls. To track a user's daily purchases, we count the number of purchase transactions completed in a day. If the user uninstalls the app on a particular day, no more observations are recorded for that user. As shown in Figure W1 (in the Web Appendix), in 67% of the days during the experiment where a purchase is completed, the users made only one purchase in those days. Therefore, we model purchases as a binary outcome variable, denoting whether the user decided to make a purchase on a particular day or not. We also track if a user uninstalled the app during the period of the experiment. Our data is right-censored because 1) the experiment ended after 6 days and 2) users could uninstall the app while the experiment is running. To address this issue, we will use a Cox Proportional Hazard survival model to model user uninstalls.

4.1.2. Consumer Characteristics

The mobile commerce app considered in this study can be downloaded and used to buy different products offered in the app across different cities in India. These cities are categorized into three tiers - tier 1, tier 2, and tier 3, based on different socio-economic factors. Socio-economic factors can play a major role in user purchasing behavior [30, 31]. Additionally, technology penetration varies by region in India [29]. In most cases, penetration is higher in city tier 1 than city tier 2 and city tier 3. The level of education is on average higher in city tier 1 as compared to other tiers. We use city tiers as proxies for how comfortable the user would be with technology and the English language. We also note the user's gender based on how they identify themselves, as a female or a male. We use gender as a control in the study to account for potential differences in online user behavior due to gender [32]. The app also collects information about the model of the phone. Using this information, we obtain the price of the phone by matching the phone model as listed in the data to the price listed on the www.mysmartprice.com database. The price of the phone controls for the economic status of a user. We include information such as gender, city tier (for the city of residence), and the phone price to account for user-level differences in our models.

4.1.3. Consumer Engagement

Consumers use the app to purchase different types of products offered in the app. We classify these product offerings into two categories based on the user involvement required in completing the purchase transaction. We define these two categories as a Low involvement Product Category, $Category_1$, and a High involvement Product category, $Category_2$. $Category_1$ involves products that require less user involvement (e.g. number of steps) in completing the transaction, e.g., paying utility bill where a user selects the service provider and approves the mode of payment (by selecting one among multiple payment gateways). $Category_2$ involves products that require higher user involvement, e.g., buying a movie ticket where a user has to select the movie, theatre, time of the show and seats. High involvement in a transaction could also be associated with higher cognitive load in completing the transaction ([33], [34]). The cognitive load associated with the amount of work needed in making a purchase in either of these product categories could result in heterogeneous effects of product variables on outcomes. To consider this product-based heterogeneous effect, we track the daily count of how many times a user converses with the chatbot in one of the products categories, $Category_1$ or $Category_2$. We call these counts $InteractionsCategory^p$ $\forall p \in \{1, 2\}$.

A user may not necessarily complete a purchase after checking a product category and can end the session before buying anything from the app. Moreover, a user may not even check a product category and just browse the app looking for different options provided in the app. To account for engagement with the app, we also consider the total number of text exchanges between a user and the chatbot in a day. We denote these text exchanges as daily $Interactions$. We also track the number of sessions by a user. We define a session as the continuous interaction with the app after the opening the app. We assume that the session has ended if we observe no user activity within the app for 10 minutes after the last action. We use $Interactions$ and number of sessions to determine the $InteractionsPerSession$ for each user.

4.1.4. App Specific Controls

Users who install and join the app during the 6-day experiment are categorized into six cohorts, based on the day of installing the app. For example, all the users who join the app on day one of the experiment are categorized in Cohort 0. Consumers in different cohorts are observed for a different number of days. For example, we could observe 6 days of data for a user who installed the app on day one of the experiment but we only observe 5 days of data for a user who installed the app on day two of the experiment (Cohort 1), provided both the users did not uninstall the app till the end of the experiment. Studies on mobile apps have shown that mobile app users are very active right after installing the app [35]. Thus it is important to track the user's age within the app. Therefore, we consider the age of the app as a control variable in our models. We define the age of the app as the number of days since the installation of the app on the user's phone. For example, if a user installs the app on a Monday, the age of the app is 0 on Monday and 1 on Tuesday.

We also note that during the period of the study, no targeted marketing campaigns were initiated by the firm on new users. There is no difference in firm contacts or app design for users in different cohorts. The firm sends notifications to all users to increase engagement with the app, however, none of these vary across users. To incorporate the effects of firm-initiated marketing contacts, we also track the number of notifications responded to by a user, $Notifications$. Summary statistics for the data collected from the mobile commerce app during the experiment is shown in Table 1.

Table 1: Data Summary for Randomized Field Experiment

	V1	**V2**
Number of Users		
Male	7442	3457
Female	1241	495
City Tier 1	5594	2659
City Tier 2	1957	987
City Tier 3	1132	306
Data Summary (Mean, Median, Max, Variance)		
Number of Daily Interactions in $Category_1$	(0.07, 0, 78, 0.64)	(0.33, 0, 61, 1.21)
Number of Daily Interactions in $Category_2$	(0.17, 0, 10, 0.17)	(0.16, 0, 11, 0.31)
Phone Price (in $)	(213.04, 188, 1066, 23402)	(199.46, 188, 1066, 21193)
Notifications Opened (in a Day)	(0.025, 0, 22, 0.29)	(0.042, 0, 13, 0.39)
Number of Sessions (by a User)	(2.13, 1, 44, 1.83)	(2.31, 1, 39, 7.47)
Interactions in a Day (by a User)	(4.41, 0, 498, 561)	(10.88, 0, 653, 1189)

User characteristics for different cohorts is shown in Table 2. The demographics data show that the sample of users in each cohort is representative of the users in the experiment (approximately two-thirds probability of installing V1 and one-third probability of installing V2). It also provides cohort-level data on user purchases and uninstalls. In Table 3, we show the total purchases and uninstalls for each cohort, based on the day of the experiment, the age of the app in the mobile phone and the version of the app.

Next, we discuss the model-free evidence to illustrate the differences in user behavior when using V1 or V2 of the mobile app considered in this study.

Table 2: Cohort Wise Data Description

	Cohort 0	**Cohort 1**	**Cohort 2**	**Cohort 3**	**Cohort 4**	**Cohort 5**
Female Identifying	261	294	284	326	304	213
Male Identifying	1803	1796	1930	1898	1999	1527
City Tier 1	1338	1396	1469	1475	1546	1162
City Tier 2	471	465	493	494	501	361
City Tier 3	255	229	252	255	256	217
Version 1	1427	1446	1489	1479	1590	1252
Version 2	637	644	725	745	713	488
Total Purchase	403	201	451	511	177	175
Total Uninstalls	132	145	252	252	189	16

4.2. Model Free Evidence

Prior to developing the econometric models to determine the contribution of incorporating a new language, we check for differences between the two populations (V1 vs V2) by conducting a t-test for the difference of means for the user metrics important to the firm. As shown in Table 4, we observe an increase in the average number of purchases and the average number of conversations per session in the treatment group (V2) as compared to the control group (V1). We do not observe a significant difference in the number of sessions. While the average number of purchases increases by 132% in the treatment group as compared to the control group, the results also indicate a 107% increase in app uninstalls in the treatment group as compared to the control group.

To test for uninstalls, we follow [36] and use a non-parametric Kaplan-Meir curve to show the difference in uninstalls between V1 and V2. Kaplan-Meier curve is a widely used non-parametric tool for comparing the survival curves of two sub-population [37]. It shows the proportion of users alive (who did not uninstall the app) based on the age of the app in their phone. Figure 5a shows the overall proportion of users who chose not to uninstall the app during the six days of the experiment. Figure 5b shows the proportion of users alive in the app depending on whether the user had V1 or V2. It shows that the probability of users not uninstalling the app is higher for users with V1 as compared to users with V2. This result falls in line with initial results in Table 4 that users with V2 are more likely to uninstall the app.

In the next section, we detail the methodology and estimation procedures to determine the impact of the introduction of the new language on two

Table 3: Cohort Wise Uninstall and Purchases

			Day of the Experiment					
Cohort	Version	Output	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
0	V1	Purchase	127	27	46	7	4	2
	V2	Purchase	82	22	27	17	9	11
	V1	Uninstall	73	5	8	8	1	0
	V2	Uninstall	27	0	5	2	2	1
1	V1	Purchase	-	73	10	3	1	3
	V2	Purchase	-	66	8	12	11	4
	V1	Uninstall	-	16	33	25	7	2
	V2	Uninstall	-	26	17	10	9	0
2	V1	Purchase	-	-	176	17	3	6
	V2	Purchase	-	-	158	25	15	12
	V1	Uninstall	-	-	63	44	14	0
	V2	Uninstall	-	-	70	45	15	0
3	V1	Purchase	-	-	-	151	11	8
	V2	Purchase	-	-	-	241	16	8
	V1	Uninstall	-	-	-	82	36	2
	V2	Uninstall	-	-	-	87	42	3
4	V1	Purchase	-	-	-	-	66	8
	V2	Purchase	-	-	-	-	92	9
	V1	Uninstall	-	-	-	-	80	5
	V2	Uninstall	-	-	-	-	99	5
5	V1	Purchase	-	-	-	-	-	61
	V2	Purchase	-	-	-	-	-	102
	V1	Uninstall	-	-	-	-	-	8
	V2	Uninstall	-	-	-	-	-	8

specific metrics: daily purchases and uninstalls.

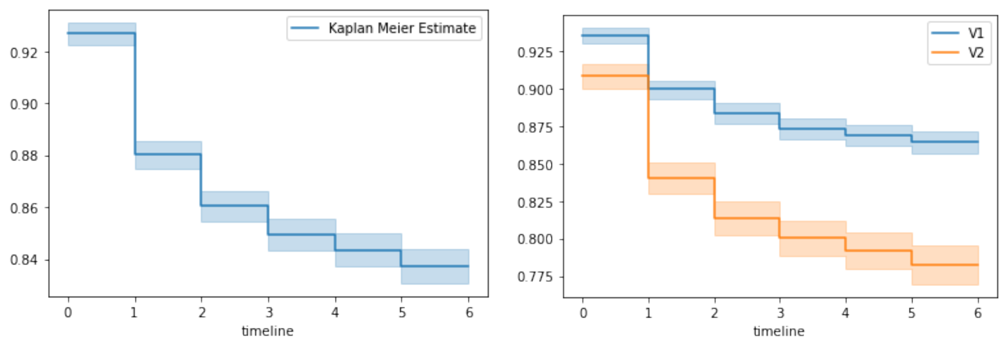
5. Model Development

We use the dummy variable $Version2_i$ to indicate if user i belonged to the control or the treatment group, where $Version2_i = 1$ if user i installed

Table 4: t-test for Consumer Metrics in the Randomized Field Experiment

	Outcome Variable	**Average**		**t-statistic**	**p-value**
		Control Group	**Treatment Group**		
1	Daily Purchases	0.025	0.058	6.45 ***	0
2	Daily Uninstalls	0.013	0.027	11.71 ***	0
3	Sessions per Consumer	1.36	2.19	23***	0
4	Interaction per session	4.41	10.88	24.85 ***	0

*p <0.1, **p <0.05, ***p <0.01



(a) Kaplan-Meier curve for all users

(b) Kaplan Meier Curve for V1 and V2

Figure 5: Kaplan Meier Curves

version V2. $Gender_i = 1$ for male-identifying user and $CityTier_{i,t} = 1$ if i lives in tier t. A user i may join the app on any day during the experiment and could also uninstall the app within the period of the study. Thus, we observe i for a total of A_i days during the experiment. We use d_i as an index for the age of the app on the user's phone, in days, after a user i installed the app on the phone. Thus, d_i can take values from $\{0, 1, ..., A_i\}$, depending on the number of days the user was in the experiment. We use Age_{i,d_i} to account for *user age* on day d_i, thereby controlling for the user's familiarity with the app. $Notifications_{i,d_i}$ is the number of notifications checked by a user i on day d_i after installing the app. $InteractionsCategory^p_{i,d_i}$ denotes the number of text exchanges with the chatbot in the product category p on day d_i. We use cohorts as a control variable to account for cohort-wise differences. $Cohort_{i,c} = 1$ if user i belongs to Cohort c. n is the number of observed data points and I is the number of users.

Next, we model the probability of purchases in subsection (5.1), followed by the model for the probability of uninstalls in subsection (5.2).

5.1. A: Purchases

Let $Purchase_{i,day_i}$ be the probability that i made a purchase on the day day_i after installing the app. Because we model the purchase decision on a given day as a binary variable, we use a logistic regression (Equation 1) to relate the version of the app and aforementioned controls to the probabilities of purchase. We also list Equation 2, that extends Equation 1 to include the cohorts indicator variable in the model.

$$
\begin{aligned}
ln\left(\frac{Purchase_{i,d_i}}{1-Purchase_{i,d_i}}\right) &= \beta_0 + \beta_1 Version2_i + \beta_2 Gender_i + \sum_{t=1}^{2}\beta_{2+t}CityTier_{i,t} + \\
&\beta_5 PhonePrice_i + \beta_6 Notifications_{i,d_i} + \sum_{p=1}^{2}\beta_{6+p}InteractionsCategory^{p}_{i,d_i} + \\
&+\beta_9 Age_{i,d_i} \quad \forall\, i \in \{1,2,..,I\}, \forall\, d_i \in \{0,1,..,A_i\}
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
ln\left(\frac{Purchase_{i,d_i}}{1-Purchase_{i,d_i}}\right) &= \beta_0 + \beta_1 Version2_i + \beta_2 Gender_i + \sum_{t=1}^{2}\beta_{2+t}CityTier_{i,t} + \\
&\beta_5 PhonePrice_i + \beta_6 Notifications_{i,d_i} + \sum_{p=1}^{2}\beta_{6+p}InteractionsCategory^{p}_{i,d_i} + \\
&+\beta_9 Age_{i,d_i} + \sum_{c=0}^{4}\beta_{10+c}Cohort_{i,c} \quad \forall\, i \in \{1,2,..,I\}, \forall\, d_i \in \{0,1,..,A_i\}
\end{aligned} \tag{2}
$$

The main effect of the introduction of local language on the increase in the probability of purchase is quantified by β_1. If β_1 is statistically significant and positive, it would indicate that language localization increases the probability of user purchases. The identification of the effect of V2 comes from comparing the purchases in the treatment group (users with V2 of the app) against the baseline purchases in the control group (users with V1 of the app).

To control for language translation effects, we also include the interaction between $Version2_i$ and the number of interactions in a given day. Finally, to control for differences in experience across less and more involved product categories, we interact $Version2_i$ with the product categories (total interaction with low and high involvement products). Details of the analysis for daily purchases is shown in Table 5.

5.2. B: Uninstalls

As noted in the experiment setup, due to the duration of the experiment, the user data are right-censored. To model user survival (a death event refers to uninstalling the mobile app from the phone), we propose a survival model that specifies "survival" probability by $S(d_i)$, i.e., the probability that a user has not uninstalled the app from the phone until d_i. Similar to [38], we use a proportional hazard model to incorporate the different control variables discussed in Section 4.1. Additionally, similar to [39], we consider cohort as a control variable to account for cohort-based heterogeneity.

To estimate the impact of different control variables and language translation (V2), we focus on the hazard rate $h(_i)$. The hazard rate is the conditional rate of uninstalling the app given that the user has kept the app on the phone for d_i days. The relationship between the survival function and the hazard rate is given in Equation 3 where $X(d_i)$ is the vector of covariates at time d_i and γ is the vector of the effect of these covariates. Hazard rate for a user i after keeping the app for d_i days is explained in Equation 4. $h_o(d_i)$ is the baseline hazard rate. We account for user-level heterogeneity using the consumer characteristics.

$$S\big(d_i|X(d_i),\gamma\big) = e^{-\sum_{v=1}^{t}\{\int_{v-1}^{v} h(d_i|X(d_i),\gamma)\,dv\}} \tag{3}$$

$$\begin{aligned} h(d_i|X(d_i),\gamma) = h_0(d_i)exp\Bigg(&\gamma_0 + \gamma_1 Version2_i + \gamma_2 Gender_i + \sum_{t=1}^{2}\gamma_{2+t}CityTier_{i,t} + \\ &\gamma_5 PhonePrice_i + \gamma_6 Notifications_{i,d_i} + \sum_{p=1}^{2}\gamma_{6+p}InteractionsCategory^{p}_{i,d_i} + \\ &+\gamma_9 Age_{i,d_i}\Bigg) \qquad \forall i \in \{1,2,..,I\} x_{d_i} \in \{0,1\}, \forall\, d_i \in \{0,1,..,A_i\} \end{aligned} \tag{4}$$

Similar to model **A**, we use different interaction terms to estimate the effect of V2 on uninstalls. The coefficient γ_1 represents the average treatment effect of V2 on user uninstalls. In addition to the model in Equation 4, we build a model in Equation 5 where we use cohorts as a control variable to account for cohort-wise differences. $Cohort_{i,c} = 1$ if user i belongs to Cohort c.

$$h(d_i|X(d_i),\gamma) = h_0(d_i)exp\Bigg(\gamma_0 + \gamma_1 Version2_i + \gamma_2 Gender_i + \sum_{t=1}^{2}\gamma_{2+t}CityTier_{i,t} + \gamma_5 PhonePrice_i + \gamma_6 Notifications_{i,d_i} + \sum_{p=1}^{2}\gamma_{6+p}InteractionsCategory^{p}_{i,d_i} + +\gamma_9 Age_{i,d_i} + \sum_{c=0}^{4}\gamma_{10+c}Cohort_{i,c}\Bigg) \quad \forall i \in \{1,2,..,I\},\ \forall\, d_i \in \{0,1,..,A_i\} \tag{5}$$

Next, we show the results for Model **A** and Model **B**.

6. Results and Discussion

6.1. Purchases

The results for the logistic regression model for Model **A**, i.e. daily purchases, using data from all the users in the experiment are shown in Table 5. The statistically significant and positive β_1 reinforces the hypothesis that a language localization feature in a chatbot-enabled mobile commerce app increases the probability that a user makes a purchase. Overall, our analysis indicates that localization could increase the odds ratio of purchase by 87% ($exp^{\beta_1} - 1$) for new users after controlling for multiple user and market characteristics. This finding supports the first hypothesis.

We further explore the potential mechanisms that explain increased purchases in the bilingual version of the app. Existing research on consumer behavior indicates that interacting with the app in the local language can increase purchases from the users. For example, the Common Sense Advisory (CSA) research report found that 76% of online shoppers, in their survey of approximately 8000 respondents, prefer to buy products with information in their native language [40]. A 2011 study from the European Commission corroborated these findings [41]. This comfort with the language during the purchase experience could provide an insight into the potential reasons as to

Table 5: Parameter Coefficients for Purchases in Model **A**

	Parameter Coefficient (Std. Error)			
	(1)	**(2)**	**(3)**	**(4)**
Intercept (β_1)	-4.51*** (0.21)	-4.56*** (0.05)	-5.30*** (0.35)	-5.40*** (0.36)
Version 2 (β_2)	0.63*** (0.19)	0.67*** (0.05)	0.62*** (0.19)	0.66*** (0.19)
Gender (β_3)	-0.55*** (0.14)	-0.55*** (0.05)	-0.52*** (0.18)	-0.52*** (0.13)
City Tier 1 (β_4)	-0.51*** (0.18)	-0.53* (0.10)	-0.48*** (0.18)	-0.50*** (0.19)
City Tier 2 (β_5)	0.07 (0.19)	0.07 (0.09)	0.11 (0.20)	0.10 (0.21)
Phone Price (β_6)	-0.12* (0.07)	-0.11*** (0.07)	-0.12* (0.07)	-0.12* (0.06)
Notifications Opened (β_7)	0.34*** (0.05)	0.44*** (0.05)	0.31*** (0.04)	0.31*** (0.04)
Interactions in Category 1 (β_8)	0.67** (0.03)	0.69** (0.04)	0.67*** (0.03)	0.69*** (0.03)
Interactions in Category 2 (β_9)	-0.05 (0.07)	0.09 (0.08)	-0.06 (0.05)	0.07 (0.07)
Age of the App (β_{10})	-0.58*** (0.07)	-0.56*** (0.07)	-0.62*** (0.07)	-0.61*** (0.07)
App Version x Category 1	-	-0.02 (0.08)	-	-0.04 (0.05)
App Version x Category 2	-	-0.65*** (0.40)	-	-0.26*** (0.08)
Cohort_0	-	-	0.87*** (0.13)	0.93*** (0.31)
Cohort_1	-	-	0.18 (0.12)	0.23 (0.33)
Cohort_2	-	-	1.01*** (0.11)	1.04*** (0.31)
Cohort_3	-	-	1.43*** (0.11)	1.45*** (0.30)
Cohort_4	-	-	0.30 (0.11)	0.37 (0.35)
n	42698	42698	42698	42698
Log Likelihood	-1443	-1432	-1413	-1403

*p <0.1, **p <0.05, ***p <0.01

why users with V2 buy more from the app as it helps them with purchasing in the local language.

Next, we observe that the low (high) involvement product category is associated with an increase (decrease) in the probability of daily purchases. Based on this observation, we explore the potential heterogeneous effects of V2 on low involvement vs high involvement categories. To isolate this effect, we interact V2 with $Category_1$ and $Category_2$. The statistically significant coefficients in Table 5 (2) and Table 5 (4) for the interaction term with $Category_2$ indicates that as users in the V2 engage in more conversations in the high involvement category, it associates with a decrease in the probability of a purchase. This result is in line with Hypothesis 2B. However, we do not find support for Hypothesis 2A as there is no statistically significant impact of the new language feature on the probability of purchase in $Category_1$, the less involved product category as users need fewer interactions to complete a purchase.

We also perform robustness checks using three different models to ensure the validity of the results (details in Section 4 in Web Appendix). First, we include the interaction between V2 and all the covariates in the model in Equation 1. In the second model, we replace the age variable with an indicator variables that corresponds to the age of the app on the user's phone (Equation W5 in Web Appendix). In the third model, we use cumulative in-app interactions as a proxy for a user's affinity for the app and use it as a control in Equation W6 (Web Appendix). The coefficients for β_1 do not change considerably for different robustness check models and stays positive and statistically significant. This shows that the results are robust and language localization in AI-assisted mobile commerce has a positive effect on increasing the probability of purchases.

6.2. Uninstalls

Next, we present the results from Cox Proportional-Hazard Model **B** in Table 6. We present the coefficients for different variables in the hazard rate. A statistically significant and positive γ_1 reinforces the result that language localization in a mobile app could increase the probability of uninstall, i.e., that users with V2 are more likely to uninstall the app than users with V1. Overall, our analysis indicates that localization could increase the hazard rate for uninstalls by 76% ($exp^{\gamma_1} - 1$) for new users. We also interact V2 with $Interactions_{i,day_i}$ and the results are shown in Table 6 (2). In Table 6 (3) and Table 6 (4), we use cohort as a control variables to account for potential cohort-level differences as shown in Equation 5.

To better understand the reasons behind the increase in the probability of uninstalls in V2, we explore the possibility that the heterogeneous effect of the product category leads to an increase in user uninstalls. We observe a potential mismatch with V2 for high involvement product $Category_2$ as shown in Table 6 (3) and Table 6 (4). The statistically significant and positive interaction of V2 and $Category_2$ indicates an increase in the hazard rate as users in V2 need to converse more in $Category_2$. This result supports Hypothesis 3 that users with vernacular version of the app are less pleased with the language localization efforts in the highly involved and more complex product interactions in $Category_2$. However, this was not true for users who purchased in the low involvement $Category_1$ as shown by a statistically significant and negative value of the coefficient of the interaction term. Results for $Category_2$ in purchases and uninstalls aligns with Hypothesis 2 that the vernacular app yields positive results for easy transactions. As noted earlier, products in $Category_2$ need more steps to complete the transactions, increasing the complexity of the transaction and the cognitive load. Given that the only difference between the two versions is the language option in the User Interface (UI), and no other changes were made in the design of the app, the results seem to indicate that the firm must consider the purchase flows and how it could vary across different languages when introducing the new language feature. This finding provides some cautionary guidance to managers who wish to include such new features that radically alter the experience within the app. The results also seem to indicate that some product categories might lend themselves more naturally to language localization than others.

We test the robustness of the findings in Table 6 using three different models (details in Section 5 in Web Appendix). In the first model, we use a user-level Cox proportional-hazard model where we keep a cumulative count of all the variables discussed in Model **B**. In the second model, we use a logistic regression model with daily data, similar to Model **A** with user uninstalls as a binary outcome. In the third model, we use logistic regression with user-level data by keeping a cumulative count of all the variables and accounting for cohort level differences in the users. Across all the three models, we observe a positive and statistically significant value of γ_1. It indicates that V2 increases the hazard rate or the probability of uninstalls, lending further support to our findings in Model **B**. Results for these robustness checks are

Table 6: Parameter Coefficients for Uninstalls in Model **B**

	Parameter Coefficient (Std. Error)			
	(1)	**(2)**	**(3)**	**(4)**
Version 2 (γ_1)	0.57*** (0.04)	0.61*** (0.06)	0.53*** (0.05)	0.56*** (0.06)
Gender (γ_2)	-0.08 (0.06)	-0.08 (0.06)	-0.06 (0.06)	-0.06 (0.06)
City Tier 1 (γ_3)	0.15 (0.1)	0.16* (0.1)	0.16* (0.1)	0.17* (0.1)
City Tier 2 (γ_4)	0.67*** (0.1)	0.67*** (0.09)	0.68*** (0.09)	0.69*** (0.09)
Phone Price (γ_5)	-0.00*** (0.0)	-0.00*** (0.0)	-0.00*** (0.0)	-0.00*** (0.0)
Notifications Opened (γ_6)	0.23*** (0.02)	0.23*** (0.02)	0.24*** (0.02)	0.23*** (0.019)
Interactions in Category 1 (γ_7)	0.05** (0.006)	0.052** (0.005)	0.04*** (0.007)	0.057*** (0.006)
Interactions in Category 2 (γ_8)	-0.05 (0.04)	-0.15 (0.07)	-0.19*** (0.05)	-0.37*** (0.07)
Age of the App (γ_9)	-0.68*** (0.03)	-0.71*** (0.03)	-0.74*** (0.03)	-0.79*** (0.03)
App Version x Category 1	-	-0.15*** (0.04)	-	-0.21*** (0.03)
App Version x Category 2	-	0.31*** (0.08)	-	0.35*** (0.08)
Cohort_0	-	-	-1.49*** (0.12)	-1.54*** (0.13)
Cohort_1	-	-	-1.07* (0.12)	-1.10*** (0.12)
Cohort_2	-	-	-0.22* (0.11)	-0.22* (0.11)
Cohort_3	-	-	0.03 (0.11)	0.02 (0.11)
Cohort_4	-	-	0.16 (0.11)	0.15 (0.11)
n	42698	42698	42698	42698
Concordance	0.789	0.79	0.804	0.805

*p < 0.1, **p < 0.05, ***p < 0.01

shown in the Web Appendix.

6.3. Control Variables

We consider multiple control variables in our models. In this section, we reconcile the main findings for different control variables in our models. While we observe increased purchases in female-identifying users in Model **A**, we do not observe any significant difference between the female-identifying and male-identifying users in the app uninstall model, at least during the initial stages of the app. Results in Table 5 indicate that purchases are higher for users in Tier 3 cities as compared against Tier 1 and Tier 2 cities. Results in Table 6 also indicate that uninstalls are lower for users in Tier 3 cities as compared to Tier 1 and Tier 2 cities. This analysis bodes well for the firm's goal of gaining access to the under-served non-English speaking population in smaller Indian cities and towns. The price of the phone is statistically insignificant for user purchases and uninstalls, thus it does not play a significant role in explaining the increase in purchases and uninstalls for the new version of the app. An increase in the number of notifications opened by a user is associated with an increase in the probability of purchase and an increase in the hazard rate for uninstalls. The former can be explained by the users' interest in learning more about the notification as they open the app after reading the notifications. However, the information may not always be what the user was expecting, hence the increase in the hazard rate by notifications opened can be explained using [42] where the authors argue that undesirable stimuli from app notifications can trigger consumer churn. While sending notifications to the users, firms must consider the usefulness of the notifications to the user as they might uninstall the app if do not find the desired information from the notifications. This could also happen when a user is annoyed by the notification as found in [43], particularly when a user is expecting correct information from the app in their vernacular. As the age of the app increases, the probability of purchase increases and the probability of uninstall decreases. This result is in line with the findings by [44], where they observe an increase in retention rate as the duration of the service increases or as users become familiar with the product.

7. Extensions

We propose two more analyses to account for heterogeneous treatment effect of V2 of the app. In the first model, we use cohort wise analysis to account for cohort - level heterogeneity in purchases. In the second model, we discuss the causal tree methodology to estimate the heterogeneous treatment effect across population.

7.1. Cohort Wise Analysis

Similar to results in Table 5, we consider data from each cohort individually and run the purchase model. We consider this approach as running six different experiments to account for cohort wise differences. We use the logistic regression model in Equation in 1 for each of the six cohorts. Results for cohort-wise logistic regression model is shown in Table 7. Note that as the cohort number increases, the maximum number of days we observe from users in that cohort decreases.

Table 7: Parameter coefficients for Cohort Analysis

	Coefficient					
Parameters	C_0	C_1	C_2	C_3	C_4	C_5
Intercept (β_0)	-6.44***	-5.77***	-5.18***	-6.21***	-5.82***	-3.71***
Version 2 (β_1)	1.14***	1.55 **	1.03**	0.88	0.64	0.13
Gender (β_2)	-0.66*	0.55	-0.86**	-1.0**	-1.29**	-1.51**
City Tier 1 (β_3)	-0.62	0.84	-1.02*	-1.03*	-0.86	-0.32
City Tier 2 (β_4)	0.09	0.52	-0.07	-1.07*	-0.35	0.46
Phone price (β_5)	0.03	0.0	0.01	-0.01*	-0.04	-0.03
Notifications (β_6)	-1.65*	-1.91**	0.46	0.26	1.46	0.94
Interactions Category 1 (β_7)	0.47***	0.43***	0.18***	0.55**	0.15	0.52*
Interactions Category 2 (β_8)	-1.06***	0.2	-0.35**	-0.69**	-1.21***	-0.13
Interactions per Day (β_9)	0.03***	0.02***	0.02***	0.03***	0.04***	0.03***
day_0 (β_{10})	0.62	0.44	0.24	1.64**	0.9	-
day_1 (β_{11})	-0.03	0.84	0.15	1.14	-	-
day_2 (β_{12})	0.95	-0.07	-0.35	-	-	-
day_3 (β_{13})	-0.31	0.14	-	-	-	-
day_4 (β_{14})	-0.39	-	-	-	-	-
n	11766	10101	8324	6298	4468	1741
Log-likelihood	-270	-114.1	-178.9	-143.1	-133	-41.9

*p <0.1, **p <0.05, ***p <0.01

Our findings in the cohort wise analysis for the effect of V2 on purchases corroborate the results in Table 5 that V2 leads to more purchases. β_1 is statistically significant and positive for cohorts zero to two. We also observe similar findings for other independent variables e.g. demographics, consumer engagement and app specific controls. Cohorts zero to two observes more data points from users. For example, potentially we can observe six data points from users in Cohort zero, five data points from Cohort one and four data points from Cohort two. From cohort three to five, β_1 is positive but not statistically significant. The result is consistent for the two product categories as increase in interaction with low (high) involvement $Category_1$ ($Category_2$) is associated with increase (decrease) in daily purchases. This indicates that there is no immediate impact of language translation on users after joining the app. Note that this results could be observed due to fewer data points per user in the newer cohorts.

Next, we discuss causal tree approach for estimating heterogeneous treatment effect of V2 on different group of users.

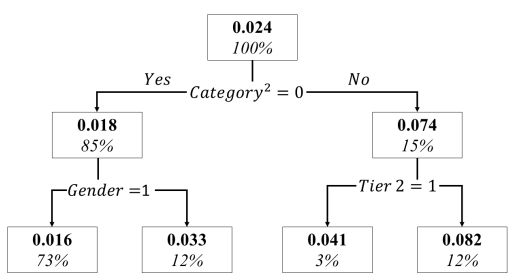
7.2. Causal Tree for Heterogeneous Treatment Effect

In Model **A** and Model **B**, we estimated average treatment effect. Estimating average treatment effects is straightforward as we use data from the randomized field experiment. However, estimation of heterogeneous treatment effects increases the complexity because it requires comparing the outcomes for matched individuals. To estimate such heterogeneous treatment effect, we use the causal trees approach proposed by [45]. A causal tree uses a tree like structure (similar to a decision tree in the Machine Learning literature) to form different subgroups by matching pairs of users in the treatment and the control group on the basis of their available covariates and comparing their respective purchases and uninstalls. The matching is performed by finding closest points using distance based metric e.g. euclidean distance. In the causal tree, closeness is defined with respect to a decision tree, and the closest points are those that fall in the same leaf. Thus, in causal tree, each leaf provides an estimate of the treatment effect for that particular subgroup.

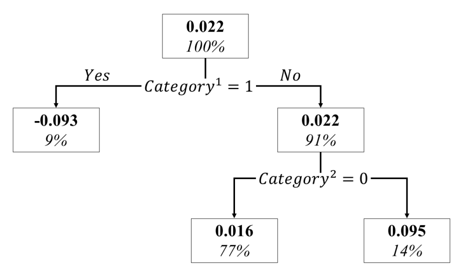
Causal trees use a non-parametric potential outcomes framework to estimate heterogeneous treatment effects. A causal tree adapts to potential outcome framework with unconfoundedness. It also adapts to the asymptotic theory for regression forests to the setting of causal inference and estimate CATE for different subgroups. Let $Y^{(1)}$ and $Y^{(0)}$ be the potential outcomes for a data point X (see [46] for a review) where $Y^{(1)}$ and $Y^{(0)}$ are respectively the response X_i would have experienced with and without the treatment (V2 in our study). The average treatment effect is defined as $E[Y_i^{(1)} - Y_i^{(0)}]$. To estimate the heterogeneous treatment effect, the causal tree finds different subgroups and it estimates the heterogeneous average conditional treatment effect (CATE) for each subgroup or leaf of the causal tree, given by $E[Y_i^{(1)} - Y_i^{(0)}|X_i]$.

A Causal tree uses the honest tree estimation method. In the honest tree estimation method, the tree is grown using one sub-sample and while the predictions at the leaves of the tree are estimated using a different sub-sample (details in [45]). In this paper, we use 50% sub-sample for building the tree and 50% sub-sample for estimating CATE for every subgroup.

We apply this approach for both purchases and uninstalls. In Figure 6a and Figure 6b, we present the first two tree partitions and corresponding CATE estimated for those partitions for V2 on daily purchases and uninstalls respectively. We present the complete causal tree in the Web Appendix in Section W7. For analysis purposes, we convert the continuous variables into binary variables for causal tree. The causal tree estimates a CATE value of 0.019 for purchases and 0.021 for uninstalls. In other words, V2 increases the probability of daily purchase by 0.019 as compared to V1. Similarly, V2 increase the probability of daily uninstall by 0.021 as compared to V1. The results from the casual trees are consistent with our findings in Model **A** and Model **B**.



(a) Causal Tree for Daily Purchases



(b) Causal Tree for Daily Uninstalls

Figure 6: Causal Tree Treatment Effects. Number in **bold** shows the estimated CATE. The number in *italic* shows the proportion of data points that are used to estimate CATE and that fall in that particular leaf or subgroup. Note that the causal tree uses honest tree approach where 50% of the data points are used to create the causal tree in this study while and the other 50% of data points are used to estimate CATE.

8. Discussion and Managerial Implications

We now reconcile the main findings from the results of our study. Theoretically, this study integrates and extends the established frameworks of the TAM, UTAUT, and CLT in the context of language localization in AI-powered chatbots. While TAM and UTAUT suggest that localization enhances perceived ease of use and usefulness, thereby positively influencing adoption and engagement, our findings reveal a more nuanced relationship. Specifically, we demonstrate that language localization, while improving purchase behavior, can also increase cognitive load in high-involvement product interactions, leading to higher user churn—a phenomenon explained through CLT. This study bridges the gap between these theories by showing how the interaction between user expectations, language switching, and task complexity mediates behavioral outcomes. Furthermore, by empirically quantifying the impact of language localization in multilingual markets, our work contributes to the growing body of literature on human-computer interaction, offering new insights into how linguistic and cultural factors influence user behavior in AI-driven platforms.

Applying these theories to practice, we run a field experiment to test the impact of language localization. The experiment allowed us, to the best degree possible, ensure that a user's behavior in the app depended only on the app itself. This assumption helped us build the models for causal inference and identification of the impact of language localization across new users. The quantification of app localization effects could be useful for product development managers who consider expanding into new markets by introducing local languages in their apps, – especially when users interact with conversational bots.

The findings reveal that new users with the new bilingual version V2 have both higher purchases and uninstalls when compared to new users with V1. While higher purchases could be expected, the finding of higher uninstalls runs counter to the conventional wisdom that argues for localization. Prior literature ([11] and [12] suggests a possible reason could be the user's language insecurity or the unexpectedness of bilingual interaction within the app that could lead to an unfavorable experience. The analysis in this study seems to indicate that the higher numbers of bilingual interactions in the high-involvement product categories could have led to greater user dissatisfaction, ultimately resulting in a higher uninstall rate, lending some support to this hypothesis.

Another possible reason for increased uninstalls with V2 could be the poor performance of the Google chatbot API for the local language. In our preliminary analysis in Table 4, we observed that the difference in the number of sessions started per user were higher for users with V2 as compared to users with V1. Additionally, the number of conversations per session for users with V2 was 146% higher than users with V1. Thus, it appears that the Google translation API (seemingly) performed satisfactorily – allowing us to rule out the effect of translation quality.

The increase in uninstalls could also be attributed to the design and use of the app when conversing in a different language. For example, individuals may have different approaches to making purchases in different languages. Multiple studies ([47], [48], [49]) have shown that people think and perceive the world differently in different languages. The findings show increased uninstalls in V2 as users interact more with the high involvement product $Category_2$, which is not the case in product $Category_1$.

So, what should a manager do – decide to develop a multi-language chatbot or not? The interaction within the chatbot's purchase process was initially optimized for the English language. The introduction of Hindi might require a new approach to the purchase flow design within the app, perhaps including fewer open-ended interactions in some product categories. Because no design alterations were made in the experiment, we cannot conclusively test for this in our data. However, recent research provides some insights into the approach firm's could employ when incorporating a new language.

[50] explore the intricate influence of language in international business (IB), integrating linguistic theories and business perspectives, with insights drawn from foreign language research in advertising. They identify three pivotal theory clusters: 'language as a symbol,' 'language in the mind,' and 'language as means of accommodation.' These clusters shed light on language's diverse meanings and mechanisms, enhancing the understanding of foreignness in emerging markets, strategic language use in foreign locales, and cross-border legitimacy in IB. Concurrently, there is an evolving shift in IB research towards a more nuanced approach to translation. This approach moves away from the traditional "*technicist*" view, which aimed for *identical* meaning between languages, this new perspective, highlighted by researchers like [51], acknowledge the complexities and nuances of language comprehension among speakers of different mother tongues.

Following this perspective, it seems that simply incorporating direct language translation innovations without accounting for other aspects of the product design and the nuances of language translation could yield negative outcomes for the firm. For example, the large increase in uninstalls should give managers pause to reconsider the purchase flow when introducing similar innovations in more intensive product categories to ensure that the in-app experience transfers seamlessly into the new language.

To maximize the benefits of language localization in the app, managers should adopt a strategy that differentiates between the complexity levels of the product categories. For low-involvement products ($Category_1$), such as utility payments and mobile phone plans, which require minimal input and follow a straightforward transaction process, the vernacular version of the app has shown positive results. Therefore, the focus for these products should be on maintaining a smooth, intuitive experience by streamlining navigation and payment flows in the local language. In contrast, high-involvement products ($Category_2$), which involve more complex interactions, such as travel bookings, hotel and local events, present a different challenge. Users engaging with these products often face higher cognitive load due to the need for additional search, decision-making, and input, which can lead to increased frustration and higher uninstall rates in the vernacular version. To mitigate these issues, managers should enhance the support available in the vernacular app for $Category_2$ by introducing guided chat prompts, tooltips, and simplified search and transaction options. Conducting A/B testing across both product categories is crucial to understanding user behavior and identifying friction points, especially for more complex transactions. A modular approach to localization could also be beneficial, allowing users to switch seamlessly between the vernacular and default language when navigating more involved purchases. This flexibility would help prevent unnecessary complexity from disrupting the user experience.

Additionally, chatbots should be trained to recognize when users encounter difficulties in the vernacular version and offer context-sensitive suggestions to streamline interactions, particularly for high-involvement products. It is important for managers to continuously monitor key performance indicators (KPIs), such as transaction completion rates, uninstalls, and time to purchase, on a category-specific basis to refine the localization strategy. By applying this data-driven, product-specific approach, managers can ensure that language localization improves user satisfaction for simpler transactions while minimizing the cognitive load and complexity associated with more involved purchases.

As most of the existing research in chatbots is done for the English language (standardized data sets and metrics for performance evaluation), firms might also benefit from considering the path to purchase when designing their apps for use in a different language. Firms introducing such innovations might consider foreshadowing the introduction of localization efforts and invest in marketing and educating the user about the innovation to reduce the unexpectedness. This effort could also help in mitigating some of the negative effects observed.

While our study provides valuable insights into the immediate impacts of language localization on user behavior, we recognize that the six-day experimental period may not fully capture the long-term dynamics of user interaction with a bilingual chatbot. It is possible that the initial increase in uninstalls observed in high-involvement product categories may diminish over time as users become more familiar with the vernacular interface and its functionalities. Over a longer period, users might adapt to the new language features, potentially leading to improved satisfaction and retention rates, especially if they perceive the localized experience as more personalized and relevant.

Conversely, if the cognitive load and complexity associated with using the vernacular version continue to present challenges, we might expect a continued or even increased uninstall rate over time. Users might grow frustrated if their expectations for a seamless, efficient interaction are not met, particularly in transactions that require extensive input or decision-making. Additionally, in the long term, the effectiveness of language localization could vary based on evolving user preferences, changing demographics, and increasing competition in the app market.

Therefore, future research should monitor these dynamics over an extended period, capturing user adaptation processes and the potential stabilization or fluctuation of key metrics like engagement, purchases, and retention. This could provide a more comprehensive understanding of how the interplay between language localization and user behavior evolves beyond the initial implementation phase.

9. Conclusions

Innovations in AI algorithms and Natural Language Processing, together with a globalized multilingual marketplace, motivate firms to invest in app localization efforts to compete more effectively in the mobile app space. In this study, we explore the impact of localization in chatbots through the introduction of a local language in a mobile app in India. We expand on the existing literature by quantifying this impact on outcome variables of interest to the firm - user purchases and uninstalls.

Leveraging a randomized field experiment, we empirically show that app localization impacts user behavior. We observe increased engagement among the treatment group in the experiment. The study also shows that these impacts could be both positive as well as negative. In our study, localization increases the odds of user purchases by up to 87%, but it also increases the hazard rate for uninstalls by about 76%. This highlights the importance of tracking multiple metrics when evaluating a feature change in an app or website based on a randomized field experiment. We note that the duration of the study could be a limitation. However, we believe the validity of the results endure because we also ensure that only the cohorts that entered the study at the same time are compared, thus maintaining a like-for-like comparison. As most apps are uninstalled within 5.8 days [3] , we believe the experiment results can provide valuable guidance to executives who may consider major technology innovations within their app.

Introducing local languages in the app as a new mode of interaction could be crucial for these apps to reach markets that were initially less accessible. Local languages may also induce a sense of connectedness and pride with the app [52]. However, the process of adding a new language in a mobile app involves changes that could affect the user experience. Additionally, because most NLP algorithms are optimized for English, words, semantics, and context, in other languages could be very different. Thus, while significant advantages exist, care must be taken in implementing the change in chatbot-assisted commerce because the localization here is more than mere translation and involves conversations with users.

We acknowledge that the focus on a single platform in India may limit the generalizability of our findings to other markets or cultural contexts. Cultural nuances, language preferences, and user behavior can vary significantly across different regions, potentially influencing the outcomes of language localization. To enhance the applicability of these insights, future research should consider extending the study to diverse platforms and markets, especially in regions with varying levels of English proficiency and differing consumer

[3] https://www.emarketer.com/content/most-apps-get-deleted-within-a-week (Accessed on 09-24-2024.)

habits. Additionally, the six-day experimental period captures immediate user responses but does not allow for an examination of long-term behavioral trends, such as user adaptation or retention. While our conclusions are grounded in objective behavioral data (e.g., purchases, uninstalls, and chatbot interactions), integrating qualitative research methods, such as user interviews or focus groups, would further deepen our understanding of user frustrations and preferences, providing richer insights into how language localization impacts user engagement and satisfaction across various contexts. This multi-faceted approach would offer a more comprehensive view of language localization's effects, ultimately guiding more effective strategies for AI-assisted platforms worldwide.

Finally, we only study the issue of language as an innovation, however, in countries like India, localization in technology products reduces the barriers to access in the new digital world. Future research could explore how such digital innovations, combined with the social, cultural, and economic fabric of the society, help lower not only the digital barriers but also lead to greater awareness and adoption across sections of society not used to transacting online.

References

[1] Markets, Markets, Conversational ai market by component (platform and services), type (iva and chatbots), technology (ml and deep learning, nlp, and asr), application, deployment mode (cloud and on-premises), vertical, and region - global forecast to 2025, `https://www.marketsandmarkets.com/Market-Reports/conversational-ai-market-49043506.html` (2020).

[2] E. Brynjolfsson, D. Rock, C. Syverson, Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics, Working Paper 24001, National Bureau of Economic Research (2017).

[3] Mckinsey, Driving impact at scale from automation and ai (2019a).

[4] A. C. Fatt, The danger of local international advertising, Journal of Marketing 31 (1967) 60–62.

[5] D. Luna, L. A. Peracchio, Moderators of language effects in advertising to bilinguals: A psycholinguistic approach, Journal of Consumer Research 28 (2) (2001) 284–295.

[6] D. Luna, L. A. Peracchio, Advertising to Bilingual Consumers: The Impact of Code-Switching on Persuasion, Journal of Consumer Research 31 (4) (2005) 760–765.

[7] S. Puntoni, B. de Langhe, S. M. J. van Osselaer, Bilingualism and the Emotional Intensity of Advertising Language, Journal of Consumer Research 35 (6) (2008) 1012–1025.

[8] J. Noriega, E. Blair, Advertising to bilinguals: Does the language of advertising influence the nature of thoughts?, Journal of Marketing 72 (5) (2008) 69–83.

[9] M. Robert, S. Hoffmann, Four positive effects of a salesperson's regional dialect in services selling, Journal of Service Research 14 (4) (2011) 460–474.

[10] E. Brynjolfsson, X. Hui, M. Liu, Does machine translation affect international trade? evidence from a large digital platform, Management Science 35 (8) (2019) 1–12.

[11] S. Koslow, P. N. Shamdasani, E. E. Touchstone, Exploring language effects in ethnic advertising: A sociolinguistic perspective, Journal of Consumer Research 20 (4) (1994) 575–585.

[12] A. Krishna, R. Ahluwalia, Language choice in advertising to bilinguals: Asymmetric effects for multinationals versus local firms, Journal of Consumer Research 35 (4) (2008) 692–705.

[13] C. Crolic, F. Thomaz, R. Hadi, A. T. Stephen, Blame the bot: Anthropomorphism and anger in customer–chatbot interactions, Journal of Marketing 86 (1) (2022) 132–148.

[14] A. Zogaj, P. M. Mähner, L. Yang, D. K. Tscheulin, It's a match! the effects of chatbot anthropomorphization and chatbot gender on consumer behavior, Journal of Business Research 155 (2023) 113412. doi:https://doi.org/10.1016/j.jbusres.2022.113412.
URL `https://www.sciencedirect.com/science/article/pii/S0148296322008773`

[15] A. Juquelier, I. Poncin, S. Hazée, Empathic chatbots: A double-edged sword in customer experiences, Journal of Business Research 188 (2025) 115074. doi:https://doi.org/10.1016/j.jbusres.2024.115074.

[16] Y. Sun, J. Chen, S. S. Sundar, Chatbot ads with a human touch: A test of anthropomorphism, interactivity, and narrativity, Journal of Business Research 172 (2024) 114403. doi:https://doi.org/10.1016/j.jbusres.2023.114403.

[17] D. Chakraborty, A. Kumar Kar, S. Patre, S. Gupta, Enhancing trust in online grocery shopping through generative ai chatbots, Journal of Business Research 180 (2024) 114737. doi:https://doi.org/10.1016/j.jbusres.2024.114737.

[18] A. P. Chaves, J. Egbert, T. Hocking, E. Doerry, M. A. Gerosa, Chatbots language design: The influence of language variation on user experience with tourist assistant chatbots 29 (2) (2022) 38. doi:10.1145/3487193.

[19] M. M. Mariani, N. Hashemi, J. Wirtz, Artificial intelligence empowered conversational agents: A systematic literature review and research agenda, Journal of Business Research 161 (2023) 113838. doi:https://doi.org/10.1016/j.jbusres.2023.113838.

[20] F. D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly 13 (3) (1989) 319–340.

[21] V. Venkatesh, F. D. Davis, A theoretical extension of the technology acceptance model: Four longitudinal field studies, Management Science 46 (2) (2000) 186–204.

[22] V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis, User acceptance of information technology: Toward a unified view, MIS Quarterly 27 (3) (2003) 425–478.

[23] J. Sweller, Cognitive load during problem solving: Effects on learning, Cognitive Science 12 (2) (1988) 257–285.

[24] F. Paas, A. Renkl, J. Sweller, Cognitive load theory and instructional design: Recent developments, Educational Psychologist 38 (1) (2003) 1–4.

[25] C. Kohler, F. A. Rohm, M. Wetzels, Return on interactivity: The impact of online agents on newcomer adjustment, Journal of Marketing 75 (2) (2011) 93–108.

[26] F. Miao, I. V. Kozlenkova, H. Wang, T. Xie, R. W. Palmatier, An emerging theory of avatar marketing, Journal of Marketing 86 (1) (2022) 67–90.

[27] R. M. Schuetzler, G. M. Grimes, J. S. Giboney, The impact of chatbot conversational skill on engagement and perceived humanness, Journal of Management Information Systems 37 (3) (2020) 875–900. doi:10.1080/07421222.2020.1790204.

[28] S. Chandra, A. Shirish, S. C. Srivastava, To be or not to be . . . human? theorizing the role of human-like competencies in conversational artificial intelligence agents, Journal of Management Information Systems 39 (4) (2022) 969–1005. doi:10.1080/07421222.2022.2127441.

[29] S. Rukmini, In india, who speaks in english, and where, https://www.livemint.com/news/india/in-india-who-speaks-in-english-and-where-1557814101428.html (2019).

[30] G. Nader, Socio-economic status and consumer behaviour, Urban Studies 6 (2) (1969) 235–245.

[31] B. Koç, M. Ceylan, The effects of social-economic status of consumers on purchasing, behaving and attitude to food products: Case study of van, turkey, British Food Journal (2012).

[32] S. Ng, M. Bharti, N. T. Faust, The Impact of Gender and Culture in Consumer Behavior, Cambridge Handbooks in Psychology, Cambridge University Press, 2020, p. 244–257.

[33] J. Sweller, P. Ayres, S. Kalyuga, Measuring cognitive load 1 (02 2011). doi:10.1007/978-1-4419-8126-46.

[34] S. Kalyuga, Cognitive load theory: How many types of load does it really need?, Educational Psychology Review 23 (1) (2011) 1–19.

[35] O. Rutz, A. Aravindakshan, O. Rubel, Measuring and forecasting mobile game app engagement, International Journal of Research in Marketing 36 (2) (2019) 185 – 199.

[36] T. Lim, Applying survival analysis for customer retention: A united states regional mobile service operator, in: 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), 2020, pp. 338–342.

[37] P. Jin, H. Haider, R. Greiner, S. Wei, G. Häubl, Using survival prediction techniques to learn consumer-specific reservation price distributions, Plos one 16 (4) (2021) e0249182.

[38] P. B. Seetharaman, P. K. Chintagunta, The proportional hazard model for purchase timing: A comparison of alternative specifications, Journal of Business Economic Statistics 21 (3) (2003) 368–382.

[39] D. A. Schweidel, P. S. Fader, E. T. Bradlow, Understanding service retention within and across cohorts using limited information, Journal of Marketing 72 (1) (2008) 82–94.

[40] D. A. DePalma, P. D. O. Mara, Can't read, won't buy – b2c (2020).

[41] N. Kelly, Speak to global customers in their own language (2017).
URL `https://hbr.org/2012/08/speak-to-global-customers-in-t`

[42] E. Ascarza, S. A. Neslin, O. Netzer, Z. Anderson, P. S. Fader, S. Gupta, B. G. Hardie, A. Lemmens, B. Libai, D. Neal, et al., In pursuit of enhanced customer retention management: Review, key issues, and future directions, Customer Needs and Solutions 5 (1) (2018) 65–81.

[43] C. Crolic, F. Thomaz, R. Hadi, A. T. Stephen, Express: Blame the bot: Anthropomorphism and anger in customer-chatbot interactions, Journal of Marketing (2021) 00222429211045687.

[44] F. F. Reichheld, T. Teal, The loyalty effect, harvard business school press, Boston, MA (1996).

[45] A. S. Wager, S., Estimation and inference of heterogeneous treatment effects using random forests, Journal of the American Statistical Association 113 (523) (2018) 1228–1242.

[46] G. W. Imbens, D. B. Rubin, Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction, Cambridge University Press, 2015. doi:10.1017/CBO9781139025751.

[47] L. Boroditsky, Does language shape thought?: Mandarin and english speakers' conceptions of time, Cognitive Psychology 43 (1) (2001) 1 – 22.

[48] S. Danziger, R. Ward, Language changes implicit associations between ethnic groups and evaluation in bilinguals, Psychology Science 21 (6) (2010) 799–800.

[49] O. Ogunnaike, Y. Dunham, M. R. Banaji, The language of implicit preferences, Journal of Experimental Social Psychology 46 (6) (2010) 999 – 1003.

[50] J. Hornikx, F. van Meurs, H. Tenzer, Foreign languages in advertising: Theoretical implications for language-related ib research, Journal of International Business Studies (2023). doi:10.1057/s41267-023-00639-6.

[51] G. Gutierrez-Huerter O, J. Moon, S. Gold, W. Chapple, Micro-processes of translation in the transfer of practices from mne headquarters to foreign subsidiaries: The role of subsidiary translators, Journal of International Business Studies 51 (3) (2020) 389–413. doi:10.1057/s41267-019-00234-8.

[52] T. Sithraputhran, Negotiating multiple identities in educational contexts: Stories of tamil heritage language users as multilingual malaysians, Ph.D. thesis, Victoria University of Wellington (2017).