Electronic Markets

Chatbot commerce - How contextual factors affect Chatbot effectiveness

Manuscript Number:	ELMA-D-22-00367R3
Full Title:	Chatbot commerce - How contextual factors affect Chatbot effectiveness
Article Type:	Research Paper
Corresponding Author:	Tuan (Kellan) Nguyen Middlesex University - Hendon Campus: Middlesex University UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Middlesex University - Hendon Campus: Middlesex University
Corresponding Author's Secondary Institution:	
First Author:	Pei-Fang Hsu
First Author Secondary Information:	
Order of Authors:	Pei-Fang Hsu
	Tuan (Kellan) Nguyen
	Chen-Ya Wang
	Pei-Ju Huang
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	The emergence of Chatbots has attracted many firms to sell their merchandise via chats and bots. Although Chatbots have received tremendous interest, little is understood about how different usage contexts affect Chatbots' effectiveness in mobile commerce. Due to differences in their nature, not all shopping contexts are suitable for Chatbots. To address this research gap, this study examines how contextual factors (i.e., intrinsic task complexity that embraces shopping task attributes and group shopping environment, and extrinsic task complexity that entails information intensity) affect user perceptions and adoption intentions of Chatbots as recommendation agents in mobile commerce. Applying the lenses of Cognitive Load Theory (CLT) and Common Ground Theory (CGT), we perform an experiment and apply quantitative analytical approaches. The results show that Chatbots are more suitable in the context of one-attribute, information-light, and group-buying tasks, whereas traditional Apps are suitable for multi-attribute, information-intensive, and single-buying scenarios. These findings make important theoretical contributions to the IT adoption literature as well as to CLT and CGT theory by contextualizing the evolving state of Chatbot commerce and providing guidelines for designing better Chatbot user experiences, thereby enhancing user perceptions and adoption intentions.

1. Introduction

Chatbots are defined as "machine conversation systems that interact with human users via natural conversational language" (Hill et al. 2015, p. 246). Users engage with Chatbots in the form of short communications through various platforms (Radziwill and Benton 2017). In recent years, firms have developed Chatbots to "chat" with customers and offer automated shopping services (Lebeuf et al. 2018). In mobile commerce, Chatbots are primarily deployed for customer interactions whereas Apps serve a much broader range of functions. For the purpose of this study, we position Chatbots and Apps as separate manifestations of recommendation agents (RA) that provide information, suggestions, and reservation functions in the context of mobile shopping. Both Chatbots and Apps provide product information and recommendations to users to ease mobile shopping. But the central difference is that traditional Apps use a graphical user interface (GUI), such as a list and a click-and-drag mode, to recommend products to users, whereas Chatbots use a conversational user interface (CUI) to deliver responses and recommendations through textual or voice queries using human language.

While the new features of Chatbots can greatly support shopping processes in mobile shopping contexts, there is mixed anecdotal evidence recommending the use of Chatbots versus traditional Apps. Given the prevalence of Chatbots in mobile commerce, it is thought that Chatbots perform better than Apps as online shopping RAs and that Chatbots might replace Apps at some point (Elimeliah 2016, Brooke 2017, Naude 2017, Arora 2019, Sajjad 2019). However, other scholars and practitioners contend that Chatbots will not replace Apps (Grover 2016, Reddy 2018, Lim et al. 2020, Kazmi 2021). Although Chatbot features are effective in some online shopping contexts, users still prefer traditional Apps in many scenarios (Grover 2016). For example, many consumers are willing to use Chatbots to book a taxi service (a relatively simple task), but hesitate to use Chatbots to book an airline ticket, which is a more complicated task that involves providing additional information, e.g., travel dates, destination, direct flight or transit, passport information and expiration date, etc. According to Sands et al. (2020), in a more dynamic service context, the nature of Chatbots may be interpreted as a lack of emotion, thus lowering the social value of service interactions. In cases where customers are angry, Chatbots have been found to have a negative effect on satisfaction (Crolic et al. 2022). These mixed findings motivate us to further investigate how contextual conditions may influence the effectiveness of Chatbots in mobile shopping contexts. Drawing on cognitive load theory (CLT) and common grounded theory (CGT), we posit that the *kind of shopping task* (task complexity) and *the type of shopping environment* (group buying) affect consumers' perceptions of Chatbots' features, and eventually consumers' final behaviors.

From a cognitive load theory (CLT) perspective, consumers' Chatbot usage might be affected by different levels of *intrinsic task complexity* (how complex a task is) and *extrinsic task complexity* (how information is presented). Booking an airline ticket is a task that requires customers to input and consider numerous pieces of complex information beforehand. In contrast, booking a taxi is a relatively simple task that requires only a few steps to perform. Equally important for affecting how consumers perceive and behave towards Chatbots is how intensively information is presented to online shoppers during task performance. Consumers have different needs for intensive or light information to perform a task. Thus, we might expect that customers' perceptions and behaviors towards Chatbots depend on the *intrinsic task complexity* and *extrinsic task complexity.* Yet, there has been unclear about which level of intrinsic and extrinsic task complexity is appropriate for Chatbots in the current state of knowledge (Cheng et al. 2022a).

From a common grounded theory (CGT) perspective, Chatbots utilization may be affected by a *group buying* mechanism, which is an important feature of Chatbots. The collaborative consumer experience is very different from traditional online shopping, which is usually conducted individually. Chatbots can be incorporated into chat groups and collaborative settings, making it possible for multiple people to interact with the bot as a group member (Lebeuf et al. 2018). Chatbots can facilitate group shopping by providing up-to-date coordination mechanisms, shared navigation support, and human language communication to assist different parties in achieving common ground, which is defined as the common knowledge and understanding held by collaborators (Pan 2019, Pichsenmeister 2017; Clark and Brennan 1991; Cheng et al., 2022b). Whether Chatbots are superior at facilitating group shopping as a collaborative performance relative to traditional navigation-based Apps is another intriguing inquiry.

To understand the shopping contexts in which Chatbots are more suitable than traditional Apps, this study manipulates two contextual factors to create different experimental contexts. *Intrinsic task complexity* includes shopping task attributes and the group shopping environment, while *extrinsic task complexity* entails information intensity. *Intrinsic* task complexity refers to the natural complexity of tasks as understood by online shoppers, while *extrinsic* task complexity refers to how information is presented to online shoppers (Sweller 2010, 2011). Specifically, this study aims to answer the following question: How do contextual factors —namely, intrinsic and extrinsic task complexity — affect users' perception and adoption of Chatbots as recommendation agents during online shopping? To investigate this research question, we conduct an experiment in which participants perform various mobile reservation tasks on both Chatbots and traditional Apps in order to understand their perception and adoption intentions of these two online shopping recommendation agents (RAs).

In contrast to prior Chatbots studies that focused on technical design, conceptual qualitative exploration, and anthropomorphic enhancement, this study contributes to our understanding of the latest Chatbots practice. This work is one of the first empirical efforts to investigate customers' perceptions and adoption intentions towards Chatbots as RA in mobile shopping environments with different contextual factors.

Although Chatbots are regarded as a trendy technology with many features assumed to outperform traditional online Apps and RAs, our results surprisingly show that Chatbots are not favorable in all circumstances. Chatbots are best suited for one-attribute, information-light, and group-buying tasks, whereas traditional Apps are more suitable for multi-attribute, information-intensive, and single-buying scenarios. These findings contribute to the IT adoption literature, CLT, and GCT by contextualizing general theories in evolving Chatbot commerce while providing practical guidelines for designing better Chatbot user experiences, thereby enhancing user perceptions and adoption intentions of Chatbots.

2. Literature Review

2.1 Prior studies on Conversational Agents - Chatbots

We conducted a literature review using various keywords, including "Chatbot", "conversational agent", and "online shopping assistant". The results (Table 1) reveal that more efforts are needed for IS society to understand Chatbot commerce better. The literature review identified two research gaps that motivate this study. First, based on our sample of reviewed papers, most studies were conducted in an exploratory fashion by qualitatively discussing the concept of Chatbots or technically examining Chatbots features. For example, scholars from the Human and Computer Interaction (HCI) realm have examined Chatbots from an architecture design perspective and focused on Chatbot dialog architecture to develop better algorithms (e.g., Hill et al. 2015; Luger and Sellen 2016; Radziwill and Benton 2017). Their findings suggest that customers may have a more satisfying experience if Chatbots have a simple design (Luger and Sellen 2016). It was also reported that Chatbot communication has not sufficiently matured, as there is still much profanity and poor vocabulary (Hill et al. 2015). Notably, no unified framework is used to describe Chatbot features comparable to those used to describe other RAs, such as traditional Apps. As Rapp et al., 2021, (p.19) indicate, "*the field lacks unified models and theories that may give explanation of fundamental aspects of the interaction experience with chatbots.*" Our study thus aims to provide a valuable framework for scholars and practitioners when evaluating Chatbots vs. Apps.

Second, our study focuses on specific contexts rather than a general setting to examine Chatbots and Apps. As Venkatesh et al. (2011) noted, most prior IS research has been criticized for implicitly assuming the independence of context and technology. In other words, prior work assumes that significant relationships between independent variables and dependent variables hold across all contexts; however, the reality is that relationships vary across contexts. Context theorizing can help IS researchers better understand why a relationship is not always significant across contexts (Venkatesh et al. 2011). Our study contributes to the literature by including contextual factors to illuminate the specific contexts (i.e., task complexity, group/single buying) in which users perceive Chatbots as being significantly better, allowing context-specific insights to emerge. Studies empirically examining users' perceptions and intentions of using Chatbots in group shopping contexts with different task complexities are limited. Our study aims to fill these gaps.

Table 1. Findings of the Literature Review

Study	Theory	Methodology	Findings
Pereira et al. (2016)	NA	Qualitative	Chatbots should have a personality, direct a conversation, pay attention to small talk, and fail like humans.
Luger and Sellen (2016)	Norman's 'gulfs of execution and evaluation'	Qualitative	The easier and more helpful the Conversational Agent, the more satisfying the customer experience.
Kumar et al. (2016)	NA	Qualitative	The authors propose a conceptual framework and propositions for adopting intelligent agent technology.
Zamora (2017)	NA	Qualitative	Users expect Chatbots to be high performing, smart, seamless, and personable.
Radziwill and Benton (2017)	NA	Qualitative	A literature review of quality issues and attributes of Chatbots' development and implementation.
Schultze and Brooks (2019)	Social presence theory	Qualitative	The authors propose a social presence model for chat-based technologies consisting of three phases: co-presence, focused interactions, and interlocking involvement obligation.
Blazevic & Sidaoui (2022)	Service, technology, and customer context	Qualitative	Search, experience, and credence influence interactions between service, technology, and customer triads during Chatbot usage.
Hill et al. (2015)	NA	Observation	People communicate with Chatbots for a longer duration and use shorter messages, more profanity, and less rich vocabulary than during human-human interactions.
Mittal et al. (2016)	NA	Performance evaluation	Chatbots have important advantages, including availability, scalability, reliability, and low cost.
Chung et al. (2018)	NA	Survey	The five features of Chatbots (interaction, entertainment, trendiness, customization, problem-solving) can enhance customer satisfaction for luxury brands

Kasilingam (2020)	- Technology acceptance model -Diffusion of innovations theory	Survey	Chatbot adoption is associated with the variables of perceived usefulness, perceived ease of use, perceived enjoyment, price consciousness, perceived risk, and personal innovativeness.
Han (2021)	Anthropomorphism	Survey	Chatbot anthropomorphism has a positive effect on customer's intention to purchase.
Cheng et al. (2022a)	Stimulus–organism–response model	Survey	Task complexity negatively moderates the relationship between Chatbot's attributes and Customer's behaviors.
Al-Natour et al. (2006)	Technological artifacts as social actors	Experiment	Customers' perceptions of similarities positively influence their evaluations of online shopping assistants.
Jenkins et al. (2007)	NA	Experiment	Users expect Chatbots to have the same tone, sensitivity, and behavior as humans but process more information.
Zhu et al. (2010)	Common ground theory	Experiment	Online communication support tools (Text/Voice) build better connections between online collaborative shoppers than navigation tools do.
Al-Natour et al. (2011)	Technological artifacts as social actors	Experiment	Perceived decision process similarity is an antecedent to enjoyment, social presence, trust, ease of use, and usefulness towards online shopping assistants.
Schuetzler et al. (2014)	Social presence theory	Experiment	Adaptive Chatbots increase perceived humanness and perceived engagement
Bickmore et al. (2016)	Conversational search	Experiment	Conversational agent-based search engine interfaces can be a good alternative to conventional Web form-based interfaces.
Araujo (2018)	Anthropomorphism	Experiment	Human-like cues (language and name) and framing can influence anthropomorphism.
Ciechanowski et al. (2019)	Theory of planned behavior	Experiment	Users are more willing to use a simpler text Chatbot than a complex one.
Lim et al. (2020)	Concept of user experience	Experiment	Rule-based Chatbots have better searchability and reliability, while natural language processing-based Chatbots have better usefulness and usability.
Borau et al. (2021)	Machine emotions and feelings	Experiment	Female chatbots are considered more human than their male counterparts.
Moriuchi (2021)	Theory of conversation	Experiment	Customers prefer interacting with an augmented reality app to a Chatbot.
Nguyen et al. (2022)	Self-determination theory	Experiment	Chatbots lead to a lower level of perceived autonomy, user satisfaction, and higher cognitive load than menu-based interface systems.
Crolic et al. (2022)	Anthropomorphism	Experiment	Chatbot anthropomorphism has a negative effect on angry customers' satisfaction.
Rapp et al. (2021)	NA	Literature review	A thorough literature review on text-based Chatbots[1].
Our study	- Cognitive load theory - Common ground theory	Experiment	Chatbots are more suitable in the context of one-attribute, information-light, and group-buying tasks, whereas traditional apps are good at multi-attribute, information-intensive, and single-buying scenarios.

2.2 Commonalities and Differences between Chatbots and Apps

Chatbots use a conversation user interface (CUI) (Nguyen et al., 2022), enabling users to interact with the software following the principles of human-to-human conversation, whereas mobile apps use a

[1] Rapp et al.'s literature review (2021) focused on text-based Chatbots while our study searched for different-based Chatbots. Rapp et al. (2021) "excluded all those papers that conducted user studies only to assess the effectiveness of a particular NLP technology or algorithm," while our study did not exclude these ones. Conference papers accounted for 57.8% in Rapp et al., (2021), while our literature review focused on journal articles.

graphical user interface (GUI) where users can follow visual hints (menu, cursors, etc.) to understand a more complex interactive system (Magalhães, 2022). On Chatbots, customers can ask for specific information and all necessary information will be provided at once, while on Apps, customers need to navigate through the different layers to find the product options they are looking for (Medina, 2020). However, in this study, both Chatbots and Apps are investigated as Recommendation Agents in the mobile commerce context, recommending products or services to mobile users. The detailed comparison between Chatbots and Apps, in the scope of this study, is given in Table 2.

Table 2. Commonalities and Differences between Chatbots and Apps

	Chatbots	**Apps**
Research context	Recommendation Agents in mobile commerce	
Online shopping goal	Product/Service recommendation and reservation	
Interface technology	Conversation User Interface (CUI)	Graphical User Interface (GUI)
Functionality	Deliver human-like responses and support	Provide detailed product-related information based on the user's search
Interaction with customers	Textual (Human language)	Navigation (Search tab, drop-down menu)
Recommendation results	A few button-like recommendations (simple and concise)	A long list of recommendations (detailed and complex)

Research gaps. Based on the review, we found that the existing Chatbot literature identifies Chatbot benefits from a piecemeal perspective. Meaningful integration of existing Chatbot features comparable to those of other RAs in the literature should be developed. Additionally, research on Chatbot features requires theoretical consideration to better explain the contexts in which the effectiveness of Chatbot features could be eroded, specifically the level of *intrinsic* and *extraneous task complexity* for which Chatbots are utilized. Our study aims to address these two gaps by synthesizing the piecemeal features and proposing a theory-guided framework to support our understanding of Chatbot design and use.

3. Research Model and Hypotheses

To examine the two conditional factors, intrinsic and extraneous task complexity, affecting Chatbot adoption and perception, we draw on CLT and CGT and propose the research model shown in Figure 1. IT adoption literature (Davis 1989; Gefen and Straub 2003) identifies two common and critical factors that affect IT adoption, *perceived ease of use (PEOU)* and *perceived usefulness (PU)*. Through our literature review of existing Chatbot studies, we identify five common and critical factors (middle of Figure 1) that are related to PEOU or PU in using Chatbot. We argue that the impacts of these factors would vary with different contexts (i.e. with different task complexity).

Ease of use. Language is the most natural interface that humans understand, as there is no need to constantly learn visual interfaces (Schlicht 2016). Compared to menu-driven navigation, language dialog provides a more natural interface that users generally prefer — especially less experienced ones (Chai et al. 2001; Radziwill and Benton 2017). Chatbots are built on a conversational interface that enables customers to make queries in their own words (Chai et al. 2001); in contrast, traditional Apps are built on a navigation interface in which customers must select, click, and repeat processes until they find the information they need. Conducting tasks using a Chatbot could be easier because customers do not need to understand the hyperlink terminologies or hierarchical menus built on websites (Chai et al. 2001). Chatbots also avoid navigation loss, in which customers feel confused about how to choose the right path to reach the desired pages (Zhang 2003). Furthermore, users do not need to download and install an App — they simply invite a bot to have a conversation and interact with it as they would with a person (Pounder 2016).

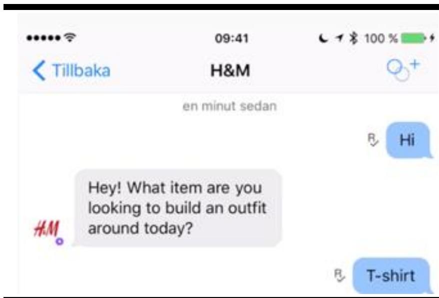
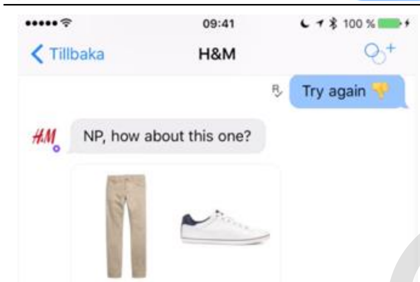
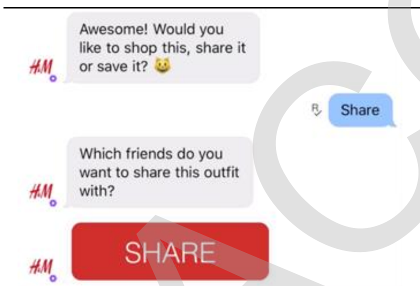
Usefulness. Chatbots can be more useful than traditional Apps in terms of time efficiency. With traditional Apps, users must specify all of their preferences upfront to obtain the recommendation output from the system; in contrast, when using a Chatbot, preferences are elicited over the length of the interaction so that efficient recommendations are provided at each stage (Chai et al. 2001; Mahmood and Ricci 2009). Additionally, when using traditional Apps, users must re-specify the search criteria to start a new search if the results are not satisfactory, whereas users of Chatbots can keep querying a previous recommendation to acquire additional information (Mittal et al. 2016; Pounder 2016). Using natural language dialog, Chatbots involve fewer clicks and reduce the time needed to obtain the relevant information (Chai et al. 2001).

Consumer engagement. Chatbots can engage consumers by detecting their identity, location, social data, and other contextual information, which, along with payment capabilities, make "conversational commerce" a substantial shopping opportunity (Pounder 2016). Because its primary function is to provide human-like responses, communication between human users and a Chatbot can bring emotions to a specific context, thereby building trust, rapport, and better communication quality (McTear et al. 2016; Io and Lee 2017). Advanced Chatbots that employ machine learning can even adapt to new contexts, new information, or new requests from customers during interactions (Radziwill and Benton 2017). Hence, Chatbots are considered to engage consumers more than other RAs during online shopping.

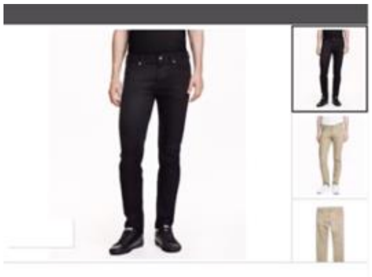
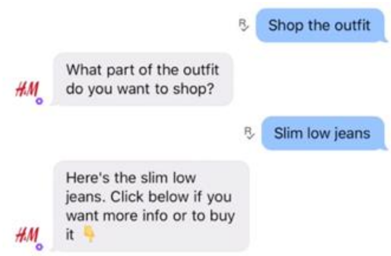
Personalized recommendations. Chatbots evolve by performing repetitive tasks to learn over time and can offer personalized recommendations (Gadiyar 2017; Pradana et al. 2017; Jenkins et al. 2007). Shopping bots use various algorithms to perform searches with keywords and present them in a consolidated and compact format that allows comparison shopping at a glance (Rowley 2000; Jenkins et al. 2007; Pradana et al. 2017). Chatbots also allow users to interact with artificial intelligence systems using their own words and their own styles to retrieve the straightforward, personalized content they want (Chai et al. 2001).

Seamless experience. Apps require users to specify preferences up front and to re-specify the search criteria every time if the results are unsatisfactory, whereas Chatbots only require users to update minimal information (Mahmood and Ricci 2009). Chatbots can also detect dead-ends in conversations and give users hints to continue an ongoing conversation (Lebeuf et al. 2018, Hsu et al. 2021). Rather than conducting separate searches using Apps, Chatbots are designed to provide a seamless experience for users at any time (Zamora 2017). To better illustrate how Chatbots recommend products to users, Table 3 provides screenshots of real examples showing *five features* of using Chatbots as RAs during mobile shopping[2].

Table 3. Examples of Five Chatbot Features

	***1. Ease-of-use.*** Chatbots use a conversational interface that enables customers to make queries directly in their own words; in contrast, traditional Apps are built on a navigation interface in which customers must select, click, and repeat processes (Chai et al. 2001).
	***2. Usefulness.*** Chatbots can keep querying a previous comment from customers to acquire additional information in real-time; when using traditional Apps, users must re-specify the search criteria to start a new search when the results are unsatisfactory (Ngai et al. 2021; Mittal et al. 2016).
	***3. Consumer engagement.*** Chatbots allow customers to invite or share their shopping results with friends to enable collaborative shopping in a group context. Chatbots provide human-like responses and bring emotions to establish trust and rapport with customers (Lebeuf et al. 2018; Li et al. 2021).

[2] We examined the literature in different domains including marketing, information systems, computer-human interactions, etc. We found that existing studies compared Chatbots and Apps using different terms for different characteristics without a united framework. Thus, during our pilot study, we asked participants to name five features of Chatbots that they think can outperform Apps. We then benchmarked these features against the extant knowledge and the Chatbot shopping examples.

	***4. Personalized recommendations*.** Chatbots use a very simple and clear button-like view and typically present only three personalized recommendations in a consolidated and compact format that allows comparison shopping at a glance; in contrast, Apps use a long list to recommend products (Gadiyar 2017; Pradana et al. 2017; Lebeuf et al. 2018).
	***5. Seamless experience.*** Chatbots can detect dead-ends in conversations and give customers hints to move an ongoing conversation towards the final goal of placing an order. There is no need to conduct a separate search using other mechanisms (Lebeuf et al. 2018; Li et al. 2021).

The IT adoption literature posits that increasing an innovation's PEOU and PU can increase user acceptance (Davis 1989; Gefen and Straub 2003). Thus, we propose that the more positive users' perceptions of the five features of a mobile shopping RA, whether Chatbots or Apps, the higher users' intentions to adopt that RA. Importantly, our model emphasizes how users' perceptions about Chatbots/Apps differ across contexts formed by two variables (i.e., *intrinsic task complexity* and *extraneous task complexity*), thereby influencing users' final adoption decisions about Chatbots/Apps during mobile shopping.

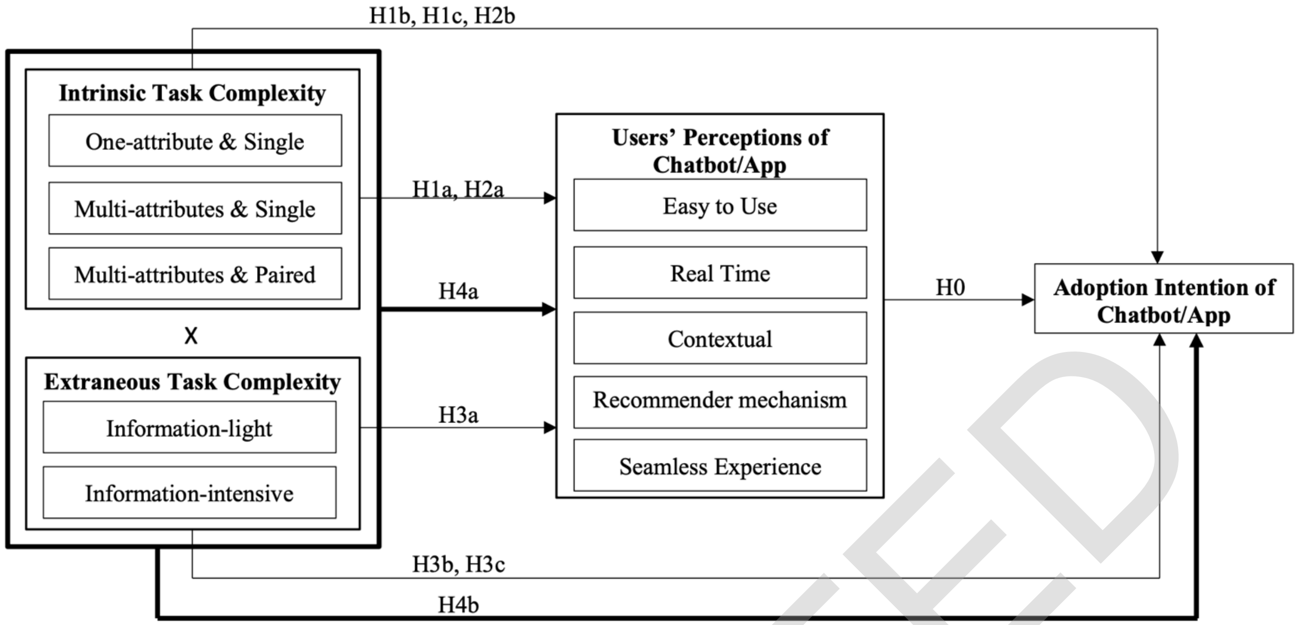


Figure 1. Conceptual framework

3.1 Cognitive Load Theory (CLT)

To understand how the contextual variables *intrinsic* and *extraneous task complexity* influence perceptions and intentions about Chatbots, we draw on CLT (Sweller 1988, 2010, 2011). According to CLT, when performing a task, humans seek to maximize the decision outcomes (e.g., accuracy, quality) while minimizing the required cognitive effort (Beach and Mitchell 1978). Accordingly, people's evaluations of a technology artifact (e.g., Chatbots or Apps) reflect their examination of how well the underlying cognitive task is supported by the artifact. If an artifact helps people achieve the desired outcome with a lighter cognitive load, they are more likely to adopt it (Al-Maskari and Sanderson, 2011). Cognitive load refers to a person's effort to process and memorize the new information needed to make decisions and can be categorized into two types: intrinsic and extraneous. *Intrinsic* cognitive load depends on the natural complexity of the information to be understood (Swell 2010, p. 124). For a given task and a given IT user's knowledge level, the intrinsic cognitive load is fixed; it can only be altered by changing the basic task or changing the user's knowledge level (Swell 2010). In contrast, *extraneous* cognitive load refers to how information is presented (Sweller 2011, p. 57) to IT users and is thus more under the control of IT/system designers. Extraneous load can be reduced by the presentation layout and instructional materials.

3.1.1 Intrinsic task complexity

In this study, *intrinsic complexity* comprises the search for product information on different numbers of product attributes (one-attribute vs. multi-attribute) and how many people input information (individual buying or group buying) on Chatbots or traditional Apps.

Product attributes. A one-attribute shopping task requires limited information and decision-making efforts (Payne 1982), whereas a multiple-attribute shopping task requires a higher cognitive load with much more attention and mental effort (Wood 1986; Speier and Morris 2003). For effective decision-making, CLT suggests that a user's cognitive load should be reasonable (Sweller 1988). Since people have limited working memory, decision-making quality and performance decrease when one's cognitive load is too high (Sweller 1988; Baddeley 1992). In other words, when the complexity level is low, the amount of information needed to be processed and the number of decisions that need to be made are limited (Payne 1982). Using an appropriate IT artifact, humans can have the sufficient working capacity to handle a low complexity task and achieve a satisfactory experience (Speier and Morris 2003). Chatbots are designed specifically to handle tasks simply and concisely using natural human language without a high learning curve; consequently, perceptions of Chatbots and intention to use could be higher for handling simple tasks. Therefore, we hypothesize

> **H1**(a): users have more positive perceptions of Chatbots when performing one-attribute tasks than multi-attribute tasks due to Chatbots' unique strength for handling simple tasks; **H1**(b): consequently, users will have higher intentions to adopt Chatbots when performing one-attribute tasks.

When tasks become complex, the cognitive load increases, thus requiring greater attention and mental effort (Speier and Morris 2003, Nguyen and Hsu 2022). When a certain level of complexity is reached, a user's performance starts to deteriorate because the task complexity has exceeded the user's capacity or the Chatbot's ability to handle the task (Wood 1986; Campbell 1988). For example, although shoppers appreciate detailed information, processing more information requires shoppers to expend more effort to reach a final purchase decision (Venkatesh et al. 2017), potentially leading to a situation where online shoppers become cognitively overwhelmed (Xu et al. 2014, Venkatesh et al. 2017, Nguyen and Hsu 2022). Therefore, when there is a very high level of complexity, users are more likely to perceive Chatbots as ineffectual and consequently have lower intentions to adopt. In contrast, Apps are superior for handling complex tasks and showing detailed information. For example, Adipat et al. (2011) report an impact of complexity on the relationship between presentation adaptations of web content and users' perceptions (ease of use, usefulness). Based on these findings, we hypothesize that:

H1(c): In contrast, Apps are better suited for multi-attribute tasks than one-attribute tasks due to their strength in handling complex information.

Group buying. Unlike earlier studies that only consider the number of product attributes when assessing intrinsic task complexity, this study also considers the *group buying* environment and investigates how this variable affects users' perceptions of Chatbots and, ultimately, adoption. ***Common Ground Theory (CGT)*** is useful for understanding group buying. It defines common ground as a shared basis of knowledge, beliefs, and suppositions among collaborators and provides a list of principles that explain how collaborators interact (Clark 1996; Romero and Markopoulos 2005). CGT argues that communication among collaborators is only effective when common ground exists, as this common knowledge makes collaborators use the same language to discuss and interpret the meaning of the messages they exchange with each other (Clark 1996). Without common ground, humans may understand and speak differently, resulting in sharing of incorrect messages (Zhu et al. 2010).

A visible common ground can enhance online shoppers' perceptions about their shopping companions' situation and preferences (i.e., one party may acknowledge the product in which the other party is interested; Kraut et al. 2003). In other words, two online shoppers may have a more effective shopping experience if they have a common ground (Zhu et al. 2010). Unlike individual shopping, group buying requires consideration of different people's preferences; the related discussion, coordination, and negotiation efforts could make reaching a final outcome during a group shopping task much more challenging. Conventional wisdom assumes that complexity is higher in a group shopping context than an individual shopping context. CGT argues that, during collaborative work, people exposed to the same working environment are more likely to recognize others' concerns and try to reach a consensus to improve productivity. Establishing a common ground (i.e., knowledge held by all partners) is thus the key to successful collaborative work (Clark and Brennan, 1991; Olson and Olson, 2000). When using Chatbots to conduct a single purchasing task, all cognitive effort must be made by one individual; in contrast, under a paired purchasing condition, two people collaborate to shop and share beliefs and knowledge to easily make a final shopping decision. Chatbots can effectively assist both parties to communicate, negotiate, and purchase using the same platform (i.e., the Chatbot). We hypothesize

> **H2** (a): users have more positive perceptions of Chatbots during group purchasing contexts than single purchasing scenario; **H2**(b): consequently, users in group purchasing scenarios will indicate higher adoption intentions towards Chatbots.

In contrast, with traditional Apps, each party uses their individual App to conduct the group task and must exchange ideas and discuss using another platform (e.g., online messaging); as a result, common ground may be more difficult to achieve. With effective communication and common ground enabled by Chatbots, paired scenarios could lead to better outcomes (Zhu et al. 2010) as shoppers efficiently help each other reach a purchase decision on the same platform without other communication or discussion, thus reducing the cognitive effort. For these reasons, we hypothesize that:

H2 (c): In contrast, when using Apps to conduct single vs. group purchasing tasks, we expect the opposite result because traditional Apps may not reduce task complexity in group shopping contexts.

3.1.2 Extraneous task complexity

CLT defines *extraneous* cognitive load as how information is presented to users (Sweller 1988, 2020, 2021). When performing an *information-intensive* task, much more information needs to be presented and displayed. According to CLT theory, tasks involving information-intensive activities challenge both humans and IT to present information in an organized way; as a result, users require more cognitive load to deal with this overload (Ortega et al. 2006). There are many presentation differences between Chatbots and Apps, so the impacts of the five common factors related to EU and EOU vary. First, Chatbots use a *button-like* view, and the number of display options is usually three to five, which is suitable for an information-light task (Lebeuf et al., 2018) for ease of use (EU). In contrast, traditional Apps use a *list-like* view and can display a very large number of options, which is suitable for displaying information-intensive tasks for their usefulness (EU). Given the relationship between innovation (Chatbots or Apps) and tasks, we hypothesize that:

H3: (a) users have more positive perceptions of Chatbots when performing information-light tasks than information-intensive tasks; (b) consequently, users performing information-light tasks will have higher adoption intentions towards Chatbots.

H3 (c): Apps are more suitable for information-intensive tasks.

3.2 Interaction effects (overall difficulty)

Because *intrinsic* and *extraneous* complexities constitute different levels of difficulty and require different levels of cognitive load, their possible interaction effects warrant further scrutiny. In this study, *overall difficulty* comprises the interaction between intrinsic and extraneous complexities. According to

CLT, humans have the sufficient working capacity to handle a low difficulty task and achieve a satisfactory performance; as tasks become more difficult, a higher cognitive load is required (Speier and Morris 2003). When reaching a certain level of difficulty, an individual's performance starts to deteriorate since the task difficulty exceeds the user's capacity to handle it (Wood 1986; Campbell 1988). Previous RA studies examining task difficulty propose that cognitive perception related to innovation may follow an inverted U-shaped path as the difficulty level increases (Xu et al. 2014). Based on CLT and the previous literature, we hypothesize that:

H4a: users' perceptions of Chatbot features follow an inverted U-shape when overall difficulty increases.

H4b: users' intentions about Chatbots follow an inverted U-shape when the overall difficulty increases.

4. Methodology

4.1 Experimental design

First, our research question is "How do contextual factors (intrinsic and extrinsic task complexity) affect users' perception and adoption of Chatbots as a recommendation agent in online shopping?", and through this research question, we would like to understand consumers' perceptions and intentions towards using the two online shopping recommendation agents. Thus, we manipulated two contextual factors intrinsic task complexity and extrinsic task complexity in the online shopping environment on either Chatbots or Apps. These intrinsic task complexity and extrinsic task complexity are suitable to be set up in the experiment rather than in other potential methods (i.e., survey). This allows the participants to truly experience different task complexity levels in the experiment rather than asking them to recall their experience with specific task complexity. Second, prior studies in the same domain also chose the experiment as the primary methodology (Al-Natour et al., 2006; Schuetzler et al., 2014, Ciechanowski et al., 2019, Borau et al., 2021, Crolic et al., 2022); thus, it motivated us to use the experiment as the main methodology to answer our research question. We performed an experiment with a 2 × 3 × 2 factorial design, with *application adoption* (Chatbots vs. Apps) and *intrinsic task complexity* (one-attribute & single vs. multi-attribute & single vs. multi-attribute & paired) as the within-subject factors and *extraneous task complexity* (information-light restaurant reservation vs. information-intensive train ticket reservation) as the between-subject factor.

To manipulate *extraneous task complexity*, we used the tasks *restaurant reservation* (information-light) and *train ticket reservation* (information-intensive) for several reasons. First, the use of different contexts will increase the generalizability and applicability of our findings. Based on our pilot study of semi-structured interviews with 20 users, users were most familiar with Chatbots and Apps used to reserve restaurants and train tickets. These two tasks were described as being highly relevant to users' daily lives compared to the other tasks suggested during the interview (e.g., healthcare, medicine, travel, film entertainment, buying clothes, etc.). Second, restaurant and train ticket reservations involve different levels of information intensity, with train ticket reservations, typically requiring more information (e.g., price, date, time, train class, and duration with a variety of output formats) than a restaurant reservation. Third, both restaurant and train ticket reservations contain some attributes that must be judged by individuals' preferences. As a result, these two tasks require discussion between two shopping partners to reach a final decision outcome in a group shopping context. As participants did not want to discuss their preferences for personal products (e.g., healthcare, medicine, etc.) with other participants, these contexts were not suitable for the group shopping experiment.

Intrinsic task complexity was manipulated by considering both the number of attributes (one vs. multiple) and the number of people inputting information (individual buying vs. group buying). Participants performing a one-attribute task must select a product under one pre-assigned attribute, such as "finding the nearest restaurant" or "booking a soon-to-depart train ticket" using a Chatbot or App. In contrast, participants performing a multi-attribute task must select products under numerous attributes, e.g., “finding a restaurant considering food categories, customer reviews, price, etc.” or “booking a train ticket with consideration of time, price, destination constraints, etc.” We further manipulate *intrinsic task complexity* by including a *group buying (vs. individual buying)* condition. In the group buying context, two participants were paired and placed in separate rooms; they were instructed to complete a group shopping task together using their own cellphones to communicate; in contrast, in the single buying context, each participant performed the task independently. Two research assistants helped conduct the experiment. One research assistant was present with each participant in a single room to provide assistance, while also unobtrusively monitoring whether the participants performed the tasks properly. Table 4 details how the different levels of task complexity and group buying were manipulated. We conducted a pre-test involving 20 participants (10 pairs) to ensure that our experimental manipulation was successful across all contexts. This manipulation is consistent with earlier studies examining complexity level and group buying (e.g., Zhu et al. 2010; Lee and Benbasat, 2011).

Table 4. Details of Task Manipulation

	Information-light (Restaurant reservation)	**Information-intensive (Train ticket reservation)**
One-attribute task (Single)	Find and book the nearest restaurant to the current location.	Find and book the latest train from Hsinchu to Taipei.
Multi-attribute task (Single)	Select and reserve a restaurant for a family reunion on Mother’s Day. Please consider the price, space, and type of restaurant that you prefer.	Select and book a train ticket to go back to your hometown. Please consider the time and type of train that you prefer.
Multi-attribute task (Group buying)	Select and reserve a restaurant together with your partner in this experiment for a college reunion in 1 month. Please consider the price, space, and type of restaurant that you both prefer.	Select and book train tickets with your partner in this experiment to return to your hometown. Please consider the time and type of train that you both prefer.

Note: We did not include *one-attribute & paired* context in this study. In the pre-test, we found that a one-attribute task (e.g., book the nearest restaurant or book the latest train ticket) is so straightforward that it generates very limited need for discussion among the pair of participants. Our interview with Chatbot commerce experts suggested that using such a seldom-occurring, unrealistic context would merely increase participants’ load when answering the questionnaire and generate limited insights.

Of existing Apps and Chatbots, the following were selected for our experimental stimuli: (1) *Taiwan Rail* App (for booking train tickets) and *EZ Table* App (for restaurant reservations), which are both accessible through two large Apps stores (iTunes App store for iOS system and Google Play store for Android system); and (2) *Taiwan Rail* Chatbot and *EZ Table* Chatbot on LINE. LINE is the most popular messaging platform in various Asian countries, including Japan, Thailand, Indonesia, etc. Chatbots on LINE offer multiple message types to enable communication between business services and users. More importantly, LINE can add Chatbots to a "group" chat, which is a unique feature suitable for our group buying context.

4.2 Participants and procedure

We recruited 174 participants (76 females and 98 males) from a public university who report being general online shoppers. The characteristics of these study participants are provided in Table 5. All participants completed pre-tests to assess their background information and online messaging-related behavior. Next, the participants completed a training session illustrating the shopping tasks using either a Chatbot or a traditional App. The formal experiment took participants about 30 minutes to complete all of the required shopping tasks and the post-questionnaires. Each participant was paid $5 USD for participation. Participants received an additional $5 USD if they joined a post-experiment interview to share more details about their Chatbot or App user experience.

Table 5. Participant Characteristics

Participants		**Count**	**%**
Gender	Male	98	56.1
	Female	76	43.6
Age	16–20	54	31.0
	21–25	114	65.5
	26–30	4	2.2
	31 and older	2	1.0
Operating system of mobile phone	Android	95	54.5
	iOS	79	45.4
Time spent on messaging Apps (hrs./day)	< 1 hr.	27	15.5
	1–3 hrs.	72	41.3
	4–5 hrs.	49	28.1
	6–7 hrs.	14	8.0
	> 7 hrs.	12	6.8

Goal of using messaging Apps (multiple)	Chat with family and friends	173	99.4
	Discuss work with colleagues	125	71.8
	Read news or catch up on new information	59	33.9
	Watch videos	35	20.1
	Receive information about product discounts or promotions	33	18.9

For the experiment, participants were first randomly assigned to one of the two extraneous task complexity groups (84 participants were assigned to the information-light restaurant reservation and 84 participants were assigned to the information-intensive train ticket reservation). Next, each participant completed three intrinsic task complexity levels (one-attribute & single (OS), multi-attribute & single (MS), multi-attribute & paired (MP)), using both a Chatbot and App. Participants compared their perceptions and intentions about using the two shopping mechanisms (Chatbots and Apps) when they performed the task at the same intrinsic task complexity level and then completed a post-experiment survey. Because each participant performed three levels of intrinsic task complexity (OS, MS, MP), three post-questionnaires were completed. To minimize potential learning effects resulting from the within-subjects design, we randomized the order of the two applications (Chatbots and Apps) as in Figure 2.

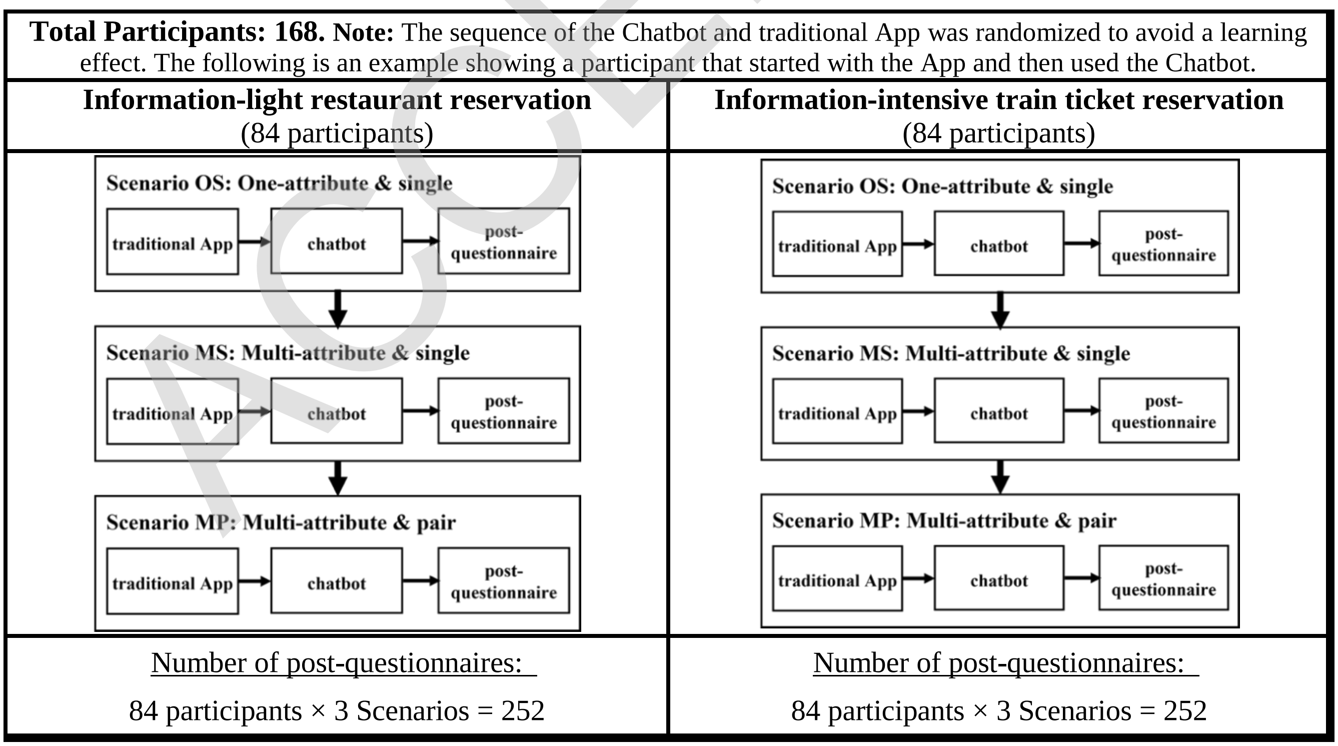


Figure 2. Experimental procedure

4.3 Measurement of user perceptions

Perceptions of the two shopping mechanisms (Chatbots, Apps) were evaluated using a five-point Likert scale, with -2 indicating "strongly disagree", and 2 indicating "strongly agree". The benefit of using this scale rather than the traditional 1–5 scale is easier interpretation. For example, participants in our study answered comparative perception questions *(e.g., "I think the Chatbot is easier to use than a traditional App when conducting this task");* if a user selects a negative value (-2, -1), it is clear that they prefer the traditional App, while a positive value (1, 2) indicates a preference for a Chatbot. *User intention* is operationalized as a categorical variable; after performing the same shopping task using the Chatbot and traditional App, participants choose their preferred shopping assistant. Users can also select the options "either" or "neither". We assessed instrument validity, including item reliability, internal consistency, and discriminant validity. All item loadings are larger than or very close to the recommended threshold of 0.7 (Appendix A), indicating that the items fit the intended constructs well (Hair et al. 2014). To examine the internal consistency of the constructs, composite reliability (CR) and Cronbach's alpha (CA) were calculated for each construct. Again, the values are larger than the recommended threshold of 0.7 (Fornell and Larcker 1981). Additionally, the discriminant validity requirement is met, as the square roots of the average variance extracted for each construct (bold diagonal elements) are greater than the inter-correlations (off-diagonal elements) between paired latent constructs (Hair et al. 2014). We also performed a collinearity test. The variance inflation factors (VIF) of all constructs are below 4.0, indicating no multicollinearity (Hair et al. 2014).

5. Results

First, we examined users' overall intentions to use for Chatbots and Apps during mobile shopping. For the assigned tasks across the different contexts, 43.7% of users preferred traditional Apps, while 25.6% preferred Chatbots. This result has an important implication: despite being a trendy technology with many features assumed to outperform traditional RAs, Chatbots may not be suitable for all mobile shopping contexts. Rather, Chatbots are only favorable in specific contexts. This result highlights the need to better understand the contexts (formed by different intrinsic and extraneous task complexity) in which users favor Chatbot commerce. Based on users' final choices, we categorize user intention as a categorical variable: Chatbot adopters (1) and non-adopters (0). We then conducted a logistic regression to understand which of the five features of Chatbots contribute most to Chatbot adoption intentions. As shown in Table 6, the three most important features are "Ease of use" ($\beta = 0.426$, $p = 0.007$), "Usefulness" ($\beta = 0.405$, $p = 0.021$), and "Personalized recommendation" ($\beta = 0.869$, $p = 0.001$).

Table 6. Logistic Regression Results

	β	S.E.	df	Sig.
Ease of use	0.426	0.159	1	**0.007****
Usefulness	0.405	0.176	1	**0.021****
Consumer engagement	-0.191	0.208	1	0.358
Personalized recommendation	0.869	0.201	1	**0.001*****
Seamless experience	0.093	0.184	1	0.615

** p < 0.1, ** p < 0.05, *** p < 0.001*

5.1 Intrinsic Task Complexity

5.1.1. Product Attributes Dimension

Next, we examined how the number of attributes directly affects users' intentions to choose Chatbots versus Apps by conducting a Chi-square test on the two complexity levels: one- vs. multi-attribute. As shown in Table 7, for one-attribute tasks, 64.3% of users chose Chatbots; for complex multi-attribute tasks, 42.9% chose Chatbots. The chi-square test results indicated a significantly higher intention to adopt Chatbots for one-attribute tasks compared to multi-attribute tasks (p = 0. 001), supporting H1b. On the other hand, significantly more users chose Apps when performing multi-attribute tasks than one-attribute tasks, due to their strengths in handling complex information, thus supporting H1c.

Table 7. Chi-square Test Results

		User Intention		Total (n)
		Chose App	Chose Chatbot	
One-attribute	Count % within simple % of total	60 (35.7%) (11.9%)	108 (64.3%) (21.4%)	168 (100%) (33.3%)
Multi-attribute	Count % within complex % of total	192 (57.1%) (38.1%)	144 (42.9%) (28.6%)	336 (100%) (66.7%)
Total (n)		252 (50%)	252 (50%)	504 (100%)
Sig. (Pearson)		**0.001*** **		

** p < 0.1, ** p < 0.05, *** p < 0.001*

We then conducted a *t*-test to understand the effect of *number of attributes* (one- vs. multi-attribute) on users' perceptions of the five Chatbot features (H1a). As shown in Table 8, users' perceptions are significantly higher for the five features when performing *one-attribute* tasks compared to *complex multi-attribute* tasks, supporting H1a. Notably, the mean differences for "usefulness" and "personalized recommendation" are the largest (mean difference = 0.442 and 0.385, respectively, p = 0.001), indicating that these two features play especially important roles in enhancing users' perceptions of Chatbots[3].

Table 8. *T*-test Results for User Perceptions (One-attribute vs. Multi-attribute Tasks)

	Attribute type	Mean	Mean Diff	p-value (Sig.)
Ease of use	One-attribute Multi-attribute	0.452 0.153	0.299	**0.006****
Usefulness	One-attribute Multi-attribute	0.536 0.094	0.442	**0.001*****
Consumer engagement	One-attribute Multi-attribute	0.386 0.108	0.278	**0.003****
Personalized recommendation	One-attribute Multi-attribute	0.148 -0.237	0.385	**0.001*****
Seamless experience	One-attribute Multi-attribute	0.597 0.249	0.348	**0.001****

** p < 0.1, ** p < 0.05, *** p < 0.001*

In the post-experiment interviews, we explored why users had better perceptions of the usefulness and recommendation features when performing *one-attribute* tasks. A common issue highlighted by users

[3] MANOVA was also used to examine the effect and yielded consistent results.

was that the Chatbot, which was designed for simplicity, only provided 2–3 products for users to choose from immediately after their search. For *one-attribute* tasks, 2–3 products were sufficient for users to make a final decision and they were significantly impressed by the Chatbot's recommendations. For example, one interviewee expressed that *"When I want to find the nearest restaurant* (one-attribute task), *I could just select one option from the three recommendations immediately provided by the Chatbot without any effort"*. This feedback supports the idea that users are pleased to have fewer options suggested by Chatbots when performing a simple task; on the other hand, users felt overwhelmed by the various product recommendations when using Apps to conduct a *one-attribute* task.

However, it is a different story when performing a multi-attribute task. The 2–3 options provided by Chatbot were insufficient for users to make a final purchase decision, as they were interested in searching other product attributes such as price, space, reviews, etc. In this context, traditional Apps led to much better user perceptions due to the mature menu-driven mechanism. In summary, we conclude that users better perceive the Chatbot's five features when performing *one-attribute* tasks versus *multi-attribute* tasks, thus supporting H1a.

5.1.2. Group Buying Dimension

We conducted a Chi-square test to understand the influence of group buying and single buying on adopting Chatbots. As shown in Table 9, for the single buying task, 36.3% of participants chose Chatbots; when performing a group purchasing task, 49.4% chose Chatbots. According to the Chi-square result (p = 0.015), there is a significant difference with higher adoption intentions for Chatbots in a *group buying* context, thus supporting H2b. On the other hand, when using Apps to perform single vs. group purchasing tasks, we see the opposite result as traditional Apps do not reduce the task complexity in a group shopping context, thus supporting H2c.

Table 9. Chi-square Test Results

		User Intention		Total (n)
		Chose App	Chose Chatbot	
Single	Count % within single % of total	107 (63.7%) (31.8%)	61 (36.3%) (18.2%)	168 (100%) (50%)
Pair	Count % within pair % of total	85 (50.6%) (25.3%)	83 (49.4%) (24.7%)	168 (100%) (50%)
Total (n)		192 (57.1%)	144 (42.9%)	336 (100%)
Sig. (Pearson)		**0.015****		

*$p < 0.1$, ** $p < 0.05$, *** $p < 0.001$*

We also conducted a *t*-test to examine user perceptions of Chatbots between single and group buying contexts to evaluate H2a. As shown in Table 10, when performing paired purchasing tasks, users perceive Chatbots to be significantly better in two features, namely "usefulness" (p = 0.009) and "seamless experience" (p = 0.061), which partially supports H2a. In the post-experiment interviews, one user commented, "*The Chatbot reduced my time copying and sharing information between my friend and me, and we can see the product options simultaneously on the Chatbot. We also spent less time searching because we can directly discuss and search in the group chatroom with the Chatbot.*" These findings support the predictions based on CGT that Chatbots can help users save time by providing a seamless shopping experience in a group purchasing context.

Table 10. *T*-test Results for User Perceptions: Single Buying vs. Group Buying (Paired)

	Group Buying	Mean	Mean Diff	Sig.
Ease of use	Single Paired	0.069 0.236	-0.167	0.217
Usefulness	single paired	-0.081 0.268	-0.349	**0.009****
Consumer engagement	single paired	0.045 0.171	-0.126	0.291
Personalized recommendation	single paired	-0.296 -0.177	-0.119	0.310
Seamless experience	single paired	0.128 0.371	-0.243	**0.061***

*$p < 0.1$, ** $p < 0.05$, *** $p < 0.001$*

5.2 Extraneous Task Complexity

To understand whether *extraneous task complexity* (information-intensive vs. information-light) has a significant impact on user intentions to choose the Chatbot, we conducted a Chi-square test. As shown in Table 11, when performing *information-light tasks* (reserving a restaurant), users have significantly higher intentions to adopt Chatbots than when performing *information-intensive tasks* (reserving train tickets) (55.6% vs. 44.4%, p = 0.013), thus supporting H3b. However, users report different adoption intentions when using Apps, thus also supporting H3c. We also conducted a t-test to evaluate H3a that "users conducting information-light tasks will have better perceptions of Chatbots than those performing information-intensive tasks". The results shown in Table 12 support this hypothesis.

Table 11. Chi-square Test Results

		User Intention		**Total**
		Chose App	Chose Chatbot	
Information-light	Count	112	140	252
	% within information-light	(44.4%)	(55.6%)	(100%)
	% of total	(22.2%)	(27.8%)	(50%)
Information-intensive	Count	140	112	252
	% within information-intensive	(55.6%)	(44.4%)	(100%)
	% of total	(27.8 %)	(22.2%)	(50%)
Total		252 (50%)	252 (50%)	504 (100%)
Sig. (Pearson)		**0.013****		

** p < 0.1, ** p < 0.05, *** p < 0.001*

Table 12. T-test Results for User Perceptions: Information-light vs. Information-intensive

	Extraneous complexity	**Mean**	**Mean Diff**	**Sig.**
Ease of use	Information-light	0.545	0.585	0.001***
	Information-intensive	-0.040		
Usefulness	Information-light	0.432	0.382	0.001***
	Information-intensive	0.050		
Consumer engagement	Information-light	0.284	0.166	0.076*
	Information-intensive	0.118		
Personalized recommendation	Information-light	0.051	0.319	0.001**
	Information-intensive	-0.268		
Seamless experience	Information-light	0.520	0.310	0.002**
	Information-intensive	0.210		

** p < 0.1, ** p < 0.05, *** p < 0.001*

The post-experiment interviews revealed further insights. One interviewee indicated that "*information overload"* and *"input flexibility"* on the user interface were the two reasons underlying their choice between Chatbots and traditional Apps. In terms of "information overload", interviewees noted that, when performing the information-light restaurant reservation task, the Chatbot clearly initiated the task with "*Hi! Do you want to find some cool restaurants?*" and what was needed from the users' side was just a confirmation. Clarity and simplicity were what attracted users to choose the Chatbot. In contrast, when using the restaurant App that provided large amounts of information (e.g., discounts, popular restaurants, latest promotions) on one screen at a time, the information overwhelmed users and they were confused about what step should be taken next. In terms of *input format flexibility*, for the information-intensive task (ticket reservation), we observed that most users encountered search errors when they tried to input complicated information to the Chatbot, such as the departure date, departure time, destination, and specific time slots. The variety of date/time formats and combinations made it difficult for the Chatbot to clearly understand the information stated by the users. As a result, users did not receive correct train timetables and ticket information and, consequently, quit using the Chatbot. Accordingly, users were not impressed by the Chatbot's features when performing the information-intensive task compared to the information-light task.

5.3 Overall Difficulty

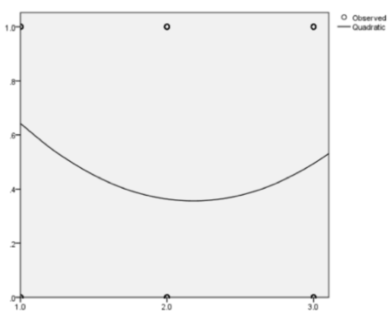
The above analyses focus on the individual effect of each contextual factor (*intrinsic task complexity* and *extraneous task complexity);* however, these variables may interact to constitute a different level of *overall difficulty* that requires a different level of cognitive load for users in each context. To explore this, we performed curvilinear regressions to regress *overall difficulty*, which includes both intrinsic and extraneous task complexity, on users' perceptions (H4a) and adoption intention (H4b) for Chatbots and Apps. As shown in Table 13, the quadratic term has a significant positive effect on users' intentions to adopt Chatbots (0.205***) and on the five features of Chatbot (0.275**, 0.483***, 0.233***, 0.282**, 0.355**), indicating a U-shaped relationship (Figure 3, left). In contrast, the quadratic term has a significant negative effect on users' intentions to adopt Apps (-0.176***), suggesting an inverted U-shaped relationship (Figure 3, right). Therefore, although the results of adoption intentions for Apps still follow CLT and show an inverted-U shape when difficulty increases, the results also show that Chatbots can better build common ground among group shopping members and ease shopping difficulty, enabling traditionally difficult group shopping tasks to be easily performed using Chatbots. The U-shape relationship differs from the predictions of H4a and H4b (vs. an inverted U-shape for traditional Apps). Thus, these results shed new light on how the latest Chatbot technology affects cognitive load and advances our understanding of CLT and CGT in these novel contexts. Furthermore, our study provides empirical evidence of nuanced differences in how users' perceived cognitive load, common ground building, and adoption intentions vary between Chatbots and Apps across contexts.

Table 13. Relationship Between Overall Difficulty and User Intentions

	DV	**IV**	Curvilinear Regression
H4a	Ease of Use	Overall Difficulty	**-1.209****
		(Std. Error)	(0.455)
		$(\text{Overall Difficulty})^2$	**0.275****
		(Std. Error)	(0.112)
	Usefulness	Overall Difficulty	**-2.064*****
		(Std. Error)	(0.449)
		$(\text{Overall Difficulty})^2$	**0.483*****
		(Std. Error)	(0.111)
	Consumer engagement	Overall Difficulty	**-1.039****
		(Std. Error)	(0.398)
		$(\text{Overall Difficulty})^2$	**0.233****
		(Std. Error)	(0.098)
	Personalized recommendation	Overall Difficulty	**-1.289****
		(Std. Error)	(0.396)
		$(\text{Overall Difficulty})^2$	**0.282****
		(Std. Error)	(0.098)
	Seamless experience	Overall Difficulty	**-1.534*****
		(Std. Error)	(0.428)
		$(\text{Overall Difficulty})^2$	**0.355****
		(Std. Error)	(0.106)
H4b	Adoption intention (Chatbot)	Overall Difficulty	**-0.896*****
		(Std. Error)	(0.186)
		$(\text{Overall Difficulty})^2$	**0.205*****
		(Std. Error)	(0.046)
	Adoption intention (Apps)	Overall Difficulty	**0.777*****
		(Std. Error)	(0.186)
		$(\text{Overall Difficulty})^2$	**-0.176*****
		(Std. Error)	(0.046)

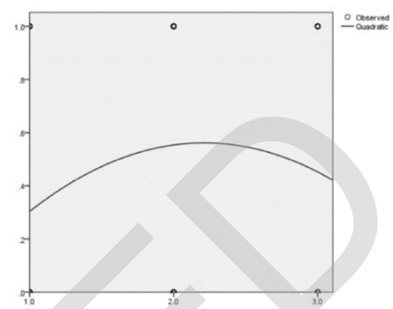
$* p < 0.1$, $** p < 0.05$, $*** p < 0.001$

Intention to adopt Chatbot



Overall Difficulty

Intention to adopt Apps



Overall Difficulty

1: One-attribute & Single, 2: Multi-attribute & Single, 3: Multi-attribute & Pair

Figure 3. Relationships between overall difficulty and adoption intention (Chatbots vs. Apps)

6. Discussion and Conclusion

Firms have rapidly adopted Chatbots to sell merchandise online. Still, little is understood about how contextual factors (intrinsic and extrinsic task complexity) affect users' perception and adoption intentions of Chatbots as recommendation agents during online shopping. By exploring this question, this study contributes to theory and practice in the following ways.

6.1. Contributions to the Chatbot literature. Although Chatbots are a trendy technology with many features assumed to outperform traditional online shopping mechanisms or RAs, our results surprisingly indicate that Chatbots are not favorable in all circumstances. Notably, different contextual factors significantly affect the effectiveness of using Chatbots during online shopping. Regarding information intensity, users who encounter information-light tasks may prefer Chatbots, but not when performing information-intensive tasks. In terms of intrinsic task complexity, users prefer a Chatbot's features when performing one-attribute tasks but not multi-attribute tasks, and users in group buying contexts, but not single buying contexts, favor Chatbots. By investigating how customers' perceptions and adoption intentions toward Chatbots vary across different contexts, we provide context-specific and user-centric insights, rather than general findings, for firms aiming to develop Chatbots commerce.

Prior Chatbot studies focus on technical design (e.g., Hill et al. 2015; Luger and Sellen 2016; Radziwill and Benton 2017), qualitative exploration (Schultze and Brooks, 2019; Blazevic and Sidaoui, 2022), and anthropomorphic enhancement (Han 2021; Araujo 2018) as shown in Table 1. This study contributes to our understanding of the latest Chatbot practice by weaving together the five Chatbot features and empirically examining customers' perceptions and adoption intentions towards Chatbots as RA in mobile shopping environments with different contextual factors. Specifically, our study addresses two research gaps identified in the Chatbot literature review. First, there is no unified framework to describe Chatbot features comparable to frameworks used to describe other RAs, such as traditional Apps. As Rapp et al., 2021, (p.19) indicate, "*the field lacks unified models and theories that may give explanation of fundamental aspects of the interaction experience with chatbots.*" Our study synthesizes various piecemeal descriptions of Chatbot features and thereby provides an integrated framework useful for scholars and practitioners interested in comparing Chatbots and Apps.

Second, our study focuses on specific, rather than generic, use contexts of Chatbots and Apps. As noted by Venkatesh et al. (2011), most prior research has been criticized for implicitly assuming the independence of context and technology, and assumes that significant relationships between independent variables and dependent variables hold across all contexts. Context theorizing can help researchers better understand why a relationship is not always significant across contexts (Venkatesh et al. 2011). For example, prior studies found that in a more dynamic service context, the nature of Chatbots may lead to a lack of emotional and social value in service interactions (Sands et al. 2020). In contexts where customers are angry, Chatbots have been found to have a negative effect on satisfaction (Crolic et al. 2022). Chatbot communication has not sufficiently matured, as there is still much profanity and poor vocabulary (Hill et al. 2015). Our study contributes to the literature by including contextual factors to illuminate the specific contexts (i.e., task complexity, group/single buying) in which users perceive Chatbots as being significantly better, allowing context-specific insights to emerge. Rapp et al.'s (2021) thorough literature review on Chatbots indicates that most existing studies examine Chatbot applications in the contexts of customer service, help desks, health care, education, etc. Yet our knowledge of the effect of applying Chatbots as an RA in mobile shopping environments with different contextual factors is limited. We remedy this knowledge gap by empirically examining two critical contextual factors.

Specifically, our study contributes in several ways by studying the contextual factors "task complexity" and "group buying" for Chatbots. For example, Cheng et al. (2022a) found that task complexity negatively moderates the relationship between Chatbot features and customer outcomes. However, in Cheng et al. (2022a), task complexity refers to intrinsic task complexity (how complex a task is) only. Our study contributes to the field of investigating both intrinsic and extrinsic task complexities, as in Campell's (1988) topology of task complexity. As for the importance of group buying, Pan (2019) discussed the messaging economy on group-messaging platforms and argued that "*recommendations of group members have a higher status and hence more influence on purchase decisions than recommendations from strangers or simply non-members.*" These arguments align with common ground theory and strengthen the need for our study to examine the group buying context. In addition, Cheng et al. (2022b) measured the collaborative performance with and without a Chatbot. The authors called on future researchers to conduct studies on collaborative performance with and without AI (i.e., Chatbots).

6.2. Contributions to Chatbot adoption frameworks. Our research model deepens our understanding of Chatbots by conceptualizing their perceived ease of use (PEOU) and perceived usefulness (PU) using five features considered particularly relevant to Chatbot commerce. Our model further includes a series of interaction effects between task complexity (which also includes group vs. individual tasks), intensity (information-light vs. information-intensive), and type of RA (Chatbot vs. traditional App) to enrich our understanding of the increasingly trend towards Chatbot-based commerce. Our paper importantly advances our understanding of the antecedents to (specifically, contextual factors) innovation acceptance, rather than focusing solely on commonly used perceptions, such as PEOU and PU, which, in the traditional literature, are believed to impact adoption intentions.

In terms of customer engagement, we provide evidence that Chatbots allow customers to invite or share their shopping results with friends to enable collaborative shopping in a group context. Chatbots can also provide human-like responses and emulate emotions in order to establish trust and rapport with customers, thereby offering new opportunities for customer engagement. To make personalized recommendations, Chatbots use a simple, button-like view and usually only present three personalized recommendations. This consolidated, compact format is very effective in mobile shopping contexts. Lastly, Chatbots can detect dead-ends in conversations and give customers hints to move the ongoing conversation towards the final goal of placing an order, enabling seamless shopping experiences. Most prior studies consider users' perceptions and intentions to use Chatbots in general, neglecting context-specific effects.

6.3. Contributions to Cognitive load theory and Common ground theory. Our comparative analysis of two online shopping RAs (Chatbots vs. Apps) strengthens our understanding of CLT and CGT theories using the latest mobile shopping context. These two theories provide cogent explanations for why the two mobile shopping agents show different effects on perceived cognitive load, perceived common ground, perceived ease of use, perceived usefulness, and adoption intention across contexts. While traditional Apps follow the inverted U-shape path depicted by CLT, the results for Chatbots deviate from this path — partially due to the enhanced common ground that Chatbots enable in group shopping contexts. Furthermore, previous studies only investigated one dimension of cognitive load, e.g., task difficulty in Lee and Benbasat (2011) and Xu et al. (2014). The present study advances our understanding of cognitive load by not only considering task difficulty (simple or complex) but also group buying (single vs. paired) and information intensity, which jointly affect cognitive load and common ground building.

6.4. Contributions to practice. Our study can also provide actionable insights to firms and managers. First, Rapp et al.'s (2021) literature review indicates that current studies mostly conduct laboratory experiments to study Chatbots, and they state that "*more effort is needed to understand how people interact with this technology in real situations*." Our study responds to this call by examining two real-world commercial Chatbots to understand their effects. We found that although Chatbots as online RAs have received much attention, users do not have distinctive preferences for Chatbots over traditional Apps in many contexts. Despite the present "Chatbot mania," firms should consider intrinsic and extrinsic task complexity when deciding whether to use or develop an App or Chatbot. The two shopping mechanisms can act complementarily to help firms optimize their business goals and performance. Second, our results provide empirical support that Chatbots generally perform best in simple, information-light, group-buying contexts. However, a more nuanced insight is that in a group shopping context, the usefulness feature (users can conveniently share information and preferences with other parties to reduce discussion and negotiation time) is the most critical driver of Chatbot adoption. We thus recommend that firms target this niche market (collaborative shopping) and increase group sales via Chatbots by further facilitating real-time environments, such as shared interfaces and collaborative decision-making tools, during conversations with Chatbots. This way, more social shopping business models will emerge in group chats.

6.5. Limitations. This study has several limitations due to technical challenges. First, we used existing Chatbots and Apps readily available to users. This choice greatly increases the external validity of our results. However, the Chatbots and Apps used in this study could have different built-in recommendation mechanisms and machine-learning algorithms. As these technical issues could not be fully controlled in our study, they should be considered in future work. Furthermore, although 76% of pairs preferred the same online shopping mechanism (Chatbot or App) for the group shopping task, 24% chose different mechanisms. As a result, although the two users were in the same shopping context, chatting with the same Chatbot, and performing the same group shopping task, they had different RA adoption intentions. It is worth further investigating these differences in the Chatbot group shopping context. In addition, most participants in this study were under the age of 30. As this age range does not cover general online shoppers, this limits generalizations of the study's findings. Lastly, as we adopted restaurant reservations and train ticket reservations as the two experimental tasks, the results may extend specifically to the food and travel industries or to online shopping in general. We encourage future researchers to deploy our theoretical framework in other industries to gain insight on the use of Chatbots and mobile Apps.

6.6. Concluding remarks. This study investigated how contextual factors affect users' perceptions and adoption intentions towards Chatbots and Apps in order to provide context-specific insights. Our results show that trendy Chatbots do not dominate across all contexts, and user preference is highly dependent on contextual features. The central takeaway of our study is that Chatbots are suitable for single-attribute, group buying, and information-light tasks, whereas traditional Apps best suited for multi-attribute, single buying, and information-intensive tasks.

References

Adipat, B., Zhang, D., and Zhou, L. 2011. “The Effects of Tree-View Based Presentation Adaptation on Mobile Web Browsing,” *MIS Quarterly* (35:1), pp. 99-121 (doi:10.2307/23043491).

Al-Maskari, A., and Sanderson, M. 2011. “The effect of user characteristics on search effectiveness in information retrieval,” *Information Processing & Management* (47:5), pp. 719-729 (doi: 10.1016/j.ipm.2011.03.002).

Al-Natour, S., Benbasat, I., and Cenfetelli, R.T. 2006. “The Role of Design Characteristics in Shaping Perceptions of Similarity: The Case of Online Shopping Assistants,” *Journal of the Association for Information Systems*, (7:12), Article 34, pp. 821-861 (doi: 10.17705/1jais.00110).

Al-Natour, S., Benbasat, I., and Cenfetelli, R.T. 2011. “The Adoption of Online Shopping Assistants: Perceived Similarity as an Antecedent to Evaluative Beliefs,” *Journal of the Association for Information Systems* (12:5), Article 2, pp.347-374 (doi: 10.17705/1jais.00267).

Araujo, T. 2018. “Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions,” *Computers in Human Behavior* (85), pp. 183-189 (doi:10.1016/j.chb.2018.03.051).

Arora, K. 2019. Will Chatbots Replace More Apps in 2019? https://www.oodlestechnologies.com/blogs/will-chatbots-replace-more-apps-in-2019/

Baddeley, A. 1992. “Working Memory: The Interface Between Memory and Cognition,” *Journal of Cognitive Neuroscience* (4:3), pp. 281-288 (doi: 10.1162/jocn.1992.4.3.281).

Beach, L.R., Mitchell, T.R. 1978. “A contingency model for the selection of decision strategies,” *Academy of Management Review* (3:3), pp. 439-449 (doi: 10.2307/257535).

Bickmore, T., Utami, D., Matsuyama, R., and Paasche-Orlow, M.K. 2016. “Improving Access to Online Health Information with Conversational Agents: A Randomized Controlled Experiment,” *Journal of medical internet research* (18:1), e1, pp. 1-12 (doi: 10.2196/jmir.5239).

Blazevic, V., & Sidaoui, K. 2022. “The TRISEC framework for optimizing conversational agent design across search, experience and credence service contexts,” *Journal of Service Management* (33:4/5), pp. 733-746.

Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. 2021. “The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI,” *Psychology & Marketing* (38:7), pp. 1052– 1068.

Brooke, C. 2017. Will Chatbots End Mobile Apps? And Why? https://www.business2community.com/mobile-apps/will-chatbots-end-mobile-apps-01853484

Campbell, D. J. 1988. “Task Complexity: A Review and Analysis,” *Academy of Management Review* (13:1), pp. 40-52 (doi: 10.2307/258353).

Chai, J., Lin, J., Zadrozny, W., Ye, Y., Stys-Budzikowska, M., Horvath, V., Kambhatla, N., and Wolf, C. 2001. “The Role of a Natural Language Conversational Interface in Online Sales: A Case Study,” *International Journal of Speech Technology* (4:3-4), pp. 285- 295 (doi: 10.1023/A:1011316909641).

Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022a), "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure", *Internet Research*

(32:2), pp. 496-517 (doi: 10.1108/INTR-08-2020-0460).

Cheng, X., Zhang, X., Yang, B., & Fu, Y. (2022b). An investigation on trust in AI-enabled collaboration: Application of AI-Driven chatbot in accommodation-based sharing economy. *Electronic Commerce Research and Applications*, 101164.

Chung, M., Ko, E., Joung, H., and Kim, S. J. 2018. "Chatbot e-service and customer satisfaction regarding luxury brands," *Journal of Business Research* (doi: 10.1016/j.jbusres.2018.10.004).

Ciechanowski, L., Przegalinska, A., Magnuski, M., and Gloor, P. 2019. "In the shades of the uncanny valley: An experimental study of human–chatbot interaction," *Future Generation Computer Systems* (92), pp. 539-548, (doi: 10.1016/j.future.2018.01.055).

Clark, H., and Brennan, S. 1991. "Grounding in Communication," 127-149 in Resnick LB, Levine JM and Teasley SD. In Lauren Resnick, Levine B., M. John, Stephanie Teasley & D. (eds.), *Perspectives on Socially Shared Cognition*, American Psychological Association, pp. 259-292.

Clark, H.H. 1996. *Using Language*, Cambridge: Cambridge University Press (doi: 10.1017/CBO9780511620539).

Crolic, C., Thomaz, F., Hadi, R., & Stephen, A. T. 2022. "Blame the bot: anthropomorphism and anger in customer-chatbot interactions," *Journal of Marketing* (86:1), pp. 132-148.

Davis, F.D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), pp. 319-340 (doi: 10.2307/249008).

Elimeliah, C. 2016. Why chatbots are replacing apps? https://venturebeat.com/business/why-chatbots-are-replacing-apps/

Fornell, C., and Larcker, D. 1981. "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *Journal of Marketing Research* (18:1), pp. 39-50 (doi: 10.1177/002224378101800104).

Gadiyar, A. R. 2017. The Chatbot Imperative: Intelligence, Personalization and Utilitarian Design. Retrieved from https://www.cognizant.com/whitepapers/the- Chatbot- imperative- intelligence- personalization- and-utilitarian- design- codex2469.pdf

Gefen, D., and Straub, D. W. 2003. "Managing User Trust in B2C e-Services." *e-Service Journal*, (2:2), pp. 7-24.

Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., Pfeiffer, B., Libai, B., Sen, S., Shi, M., and Verlegh, P. 2005. "The Firm's Management of Social Interactions," *Marketing Letters* (16:3), pp. 415- 428 (doi:10.1007/s11002- 005- 5902- 4).

Grover, D. 2016. Bots won't replace apps. Better apps will replace apps. http://dangrover.com/blog/2016/04/20/bots-wont-replace-apps.html.

Hair Jr, J.F., Black, W.C, Babin, B.J., and Anderson, R.E. 2014. *Multivariate data analysis*, Harlow, Essex: Pearson Education Limited.

Han, M. C. 2021. "The impact of anthropomorphism on consumers' purchase decision in chatbot commerce," *Journal of Internet Commerce* (20:1), pp. 46-65.

Hill, J. Ford, W.R., and Farreras, I.G. 2015. "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–Chatbot conversations," *Computers in Human Behavior* (49), pp. 245-250 (doi: 10.1016/j.chb.2015.02.026).

Hsu, P.-F., Nguyen, T. (K)., and Huang, J. 2021. “Value co-creation and co-destruction in self-service technology: A customer’s perspective,” *Electronic Commerce Research and Applications* (46), 101029 (doi: 10.1016/j.elerap.2021.101029).

Hu, J., Huhmann, B.A., and Hyman, M.R. 2007. “The Relationship between task complexity and information search: The role of self- efficacy,” *Psychology and Marketing* (24:3), pp. 253-270 (doi: 10.1002/mar.20160).

Io, H.N., and Lee, C.B. 2017. “Chatbots and conversational agents: A bibliometric analysis,” IEEE *International Conference on Industrial Engineering and Engineering Management*, Singapore, pp. 215-219 (doi: 10.1109/IEEM.2017.8289883).

Jenkins, M.C., Churchill, R., Cox, S., and Smith, D. 2007. “Analysis of User Interaction with Service Oriented Chatbot Systems,” in J. A. Jacko (Ed.), *Human-Computer Interaction, HCI Intelligent Multimodal Interaction Environments*, pp. 76-83 (doi: 10.1007/978-3-540-73110-8_9).

Kasilingam, D. L. 2020. “Understanding the attitude and intention to use smartphone chatbots for shopping,” *Technology in Society* (62), 101280.

Kazmi, R. 2021. Will Bots Really Replace Apps? https://www.koombea.com/blog/will-bots-really-replace-apps/

Kraut, R. E., Fussell, S.R., Siegel, J. 2003. “Visual information as a conversational resource in collaborative physical tasks,” *Journal Human-Computer Interaction* (18:1), pp. 13–49 (doi: 10.1207/S15327051HCI1812_2).

Kumar, V., Dixit, A., Javalgi, R.G., and Dass, M. 2016. “Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing*,” Journal of the Academy of Marketing Science* (44:1), pp. 24-45 (doi: 10.1007/s11747-015-0426-9).

Lebeuf, C., Storey, M., and Zagalsky, A. 2018. "Software Bots," *IEEE Software* (35:1), pp. 18-23 (doi:10.1109/MS.2017.4541027).

Lee, Y., and Benbasat, I. 2011. “Research Note: The Influence of Trade-off Difficulty Caused by Preference Elicitation Methods on User Acceptance of Recommendation Agents Across Loss and Gain Conditions,” *Information Systems Research* (22:4), pp. 867-884 (doi: 10.1287/isre.1100.0334).

Li, L., Lee, K.Y., Emokpae, E., and Yang, S-B. 2021. “What makes you continuously use chatbot services? Evidence from chinese online travel agencies,” *Electronic Markets* (31), pp. 575-599 (doi: 10.1007/s12525-020-00454-z).

Lim, Y., Lim, J., & Cho, N. 2020. “An Experimental Comparison of the Usability of Rule-based and Natural Language Processing-based Chatbots,” *Asia Pacific Journal of Information Systems* (30:4), pp. 832-846.

Luger, E., and Sellen, A. 2016. “Like Having a Really Bad PA: The Gulf between User Expectation and Experience of Conversational Agents,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 5286-5297 (doi: 10.1145/2858036.2858288).

Magalhães, R. 2022. Conversational User Interface: The Ultimate Breakdown. https://landbot.io/blog/conversational-interfaces-explained

Mahmood, T., and Ricci, F. 2009. “Improving recommender systems with adaptive conversational strategies,” in *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pp. 73-82 (doi:

10.1145/1557914.1557930).

Makadia, M. 2017. Will Chatbots end the Mobile Apps? And When? https://www.business2community.com/mobile-apps/will-chatbots-end-mobile-apps-01853484.

McTear, M., Callejas, Z., and Griol, D. 2016. *The Conversational Interface: Talking to Smart Devices*. Springer International Publishing Switzerland (doi: 10.1007/978-3-319-32967-3).

Medina, G.G. 2020. From Graphical Interface to the Conversational Interface. https://uxdesign.cc/from-the-graphical-interface-to-the-conversational-interface-4d7ffbd05be1

Mittal, A., Agrawal, A., Chouksey, A., Shriwas, R., and Agrawal, S. 2016. "A Comparative Study of Chatbots and Humans," *International Journal of Advanced Research in Computer and Communication Engineering* (5:3), pp. 1055-1057 (doi:10.17148/IJARCCE.2016.53253).

Moriuchi, E. 2019. "Okay, Google!: An empirical study on voice assistants on consumer engagement and loyalty." *Psychology & Marketing* (36:5), pp. 489-501.

Moriuchi, E., Landers, V.M, Colton, D & Hair, N. 2021. "Engagement with chatbots versus augmented reality interactive technology in e-commerce," *Journal of Strategic Marketing* (29:5), pp. 375-389.

Naude, D. 2017. Chatbots won't replace people. They'll replace apps. https://medium.com/dawidnaude/chatbots-wont-replace-people-they-ll-replace-apps-8b13c555d67b

Ngai, E. W., Lee, M. C., Luo, M., Chan, P. S., & Liang, T. 2021. "An intelligent knowledge-based chatbot for customer service," *Electronic Commerce Research and Applications* (50), 101098.

Nguyen, Q.N., Sidorova, A., and Torres, R. 2022. "User interactions with chatbot interfaces vs. Menu-based interfaces: An empirical study," *Computers in Human Behavior* (128), 107093.

Nguyen, T.(K.), and Hsu, P.-F. 2022. "More Personalized, More Useful? Reinvestigating Recommendation Mechanisms in Ecommerce," *International Journal of Electronic Commerce* (26:1), pp. 90-122 (doi: 10.1080/10864415.2021.2010006)

Olson, G.M. and Olson, J.S. 2000. "Distance matters," *Human-Computer Interaction* (5:2/3), pp. 139–178. (doi: 10.1207/S15327051HCI1523_4)

Ortega, B.H., Martinez, J.J, Hoyos, M.J.M.D. 2006. "Analysis of the moderating effect of industry on online behavior," *Online Information Review* (30:6), pp. 681-698 (doi: 10.1108/14684520610716162).

Pan, J. 2019. Are Private Group Chats The Future of Messaging Marketing? https://landbot.io/blog/group-chats-marketing.

Payne, J. 1982. "Contingent Decision Behavior," *Psychological Bulletin* (92:2), pp. 382-402 (doi: 10.1037/0033-2909.92.2.382).

Pereira, M. J., Coheur, L., Fialho, P., and Ribeiro, R. 2016. "Chatbots' Greetings to Human- Computer Communication," *arXiv*:1609.06479v1.

Pichsenmeister, D. 2017. Bot trends 2017. https://medium.com/@pichsenmeister/bot-trends-2017-78f9ecd6317f.

Pounder J, 2016. Huddle — humanity in the machine. https://www.mindshareworld.com/sites/default/files/mindshare_huddle_humanity_machine_2016_0.pdf.

Pradana, A., Sing, G.O., and Kumar, Y. J. 2017. “SamBot-Intelligent Conversational Bot for Interactive Marketing with Consumer-centric Approach,” *International Journal of Computer Information Systems and Industrial Management Applications* (6), pp. 265-275 (ISSN 2150-7988).

Qiu, L. and Benbasat, I. 2019. Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems,” *Journal of Management Information Systems* (25:4), pp. 145-181.

Radziwill, N. M. and Benton, M. C. 2017. “Evaluating Quality of Chatbots and Intelligent Conversational Agents,” *Software Quality Professional* (arXiv:1704.04579v1).

Rapp, A., Lorenzo, C., & Arianna, B. 2021. "The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots." *International Journal of Human-Computer Studies* (151), 102630.

Reddy, R. 2018. Chatbots Vs Apps: What Does Your Business Need? https://botcore.ai/blog/chatbots-vs-apps-what-does-your-business-need/.

Romero, N., and Markopoulos, P. 2005. “Common Ground as Privacy Mechanism in the use of Awareness Systems”, in *Proceedings of the 2005 IFIP TC13 international conference on Human-Computer Interaction*, pp.1006-1009.

Rowley, J. 2000. “Product searching with shopping bots,” *Internet Research* (10:3), pp. 203- 214 (doi: 10.1108/10662240010331957).

Sajjad, F. 2019. Will Chatbots Kill Apps and Websites? https://cyfuture.com/blog/will-chatbots-kill-apps-and-websites/

Sands, S., Ferraro, C., Campbell, C., & Tsao, H. Y. 2020. Managing the human–chatbot divide: how service scripts influence service experience. Journal of Service Management (10.1108/JOSM-06-2019-0203).

Schlicht, M. 2016. How Bots Will Completely Kill Websites and Mobile Apps? https://chatbotsmagazine.com/how-bots-will-completely-kill-websites-and-mobile-apps-656db8e6fc03.

Schuetzler, R., Grimes, M., Giboney, J., and Buckman, J. 2014. “Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment,” in *Proceedings of the 35th International Conference on Information Systems*, Auckland.

Schultze, U., and Brooks, J.A.M. 2019. “An interactional view of social presence: Making the virtual other “real”,” *Information Systems Journal* (29:3), pp. 707 - 737.

Speier, C., and Morris, M.G. 2003. “The Influence of Query Interface Design on Decision-Making Performance,” *MIS Quarterly* (27:3), pp. 397-423 (doi: 10.2307/30036539).

Sweller, J. 1988. “Cognitive Load During Problem Solving: Effects on Learning,” *Cognitive Science* (12:2), pp. 257-285 (doi: 10.1016/0364-0213(88)90023-7).

Sweller, J. 2010. “Element interactivity and intrinsic, extraneous, and germane cognitive load,” *Educational psychology review* (22:2), pp. 123-138.

Sweller, J. 2011. “Cognitive load theory” In *Psychology of learning and motivation* (55), pp. 37-76, Academic Press.

Venkatesh, V., Aloysius, J., Hoehle, H., and Burton, S. 2017. "Design and Evaluation of Auto-ID Enabled Shopping Assistance Artifacts in Customers' Mobile Phones: Two Retail Store Laboratory Experiments," *MIS Quarterly* (41:1), pp.83-113 (doi: 10.25300/MISQ/2017/41.1.05).

Venkatesh, V., Thong, J.Y.L., Chan, F.K.Y., Hu, P.J., and Brown, S.A. 2011. "Extending the two-stage information systems continuance model: Incorporating UTAUT predictors and the role of context," *Information Systems Journal* (21:6), pp. 527–555 (doi: 10.1111/j.1365-2575.2011.00373.x).

Wood, R. E. 1986. "Task complexity: Definition of the construct," *Organizational Behavior and Human Decision Processes* (37:1), pp. 60–82 (doi: 10.1016/0749-5978(86)90044-0).

Xu, D.J., Benbasat, I., and Cenfetelli, R.T. 2014. "The Nature and Consequences of Trade-Off Transparency in the Context of Recommendation Agents," *MIS Quarterly* (38:2), pp. 379-406 (doi: 10.25300/MISQ/2014/38.2.03).

Zamora, J. 2017. "Rise of the Chatbots: Finding a Place for Artificial Intelligence in India and US," in *Proceedings of the 22nd International Conference on Intelligent User Interfaces Companion*, pp.109-112.

Zhang, D. 2003. "Delivery of personalized and adaptive content to mobile devices: a framework and enabling technology," *Communications of the Association for Information Systems* (12), pp. 183-202.

Zhu, L., Benbasat, I., and Jiang, Z. 2010. "Let's Shop Online Together: An Empirical Investigation of Collaborative Online Shopping Support," *Information Systems Research* (21:4), pp. 872-891 (doi:10.1287/isre.1080.0218).