



# How do Artificial Intelligence Chatbots Affect Customer Purchase? Uncovering the Dual Pathways of Anthropomorphism on Service Evaluation

Yang Li<sup>1</sup> · Zhenghua Gan<sup>1</sup> · Bowen Zheng<sup>2</sup> 

Accepted: 23 September 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Although chatbots are increasingly deployed in customer service to reduce the burden of human labor and sometimes replace human employees in online shopping, there remains the challenge of ensuring consumers' service evaluation and purchase decisions after chatbot service. Anthropomorphism, referring to human-like traits exhibited by non-human entities, is considered a key principle to facilitate customers' positive evaluation of chatbot service and purchase decisions. However, equipping chatbots with anthropomorphism should be planned and rolled out cautiously because it could be both advantages to building customer trust and disadvantages for increasing customer overload. To understand how customers process and react to chatbot anthropomorphism, this study applied Wixom and Todd's model and social information processing theory which guide this study to examine how object-based social beliefs (i.e., chatbot warmth and chatbot competence) of anthropomorphic chatbot influence service evaluation and customer purchase by generating behavioral beliefs (i.e., trust in chatbot and chatbot overload). The research model was examined with a "lab-in-the-field" experiment of 212 samples and two scenario-based experiments of 124 samples and 232 samples. The results showed that chatbot warmth and competence had significant effects on trust in chatbot and chatbot overload. Trust in chatbot and chatbot overload further significantly impact service evaluation and then customer purchase. Implications for theory and practice are discussed.

**Keywords** Chatbot · Anthropomorphism · Trust · ICT-related overload · Service evaluation

## 1 Introduction

With the surge of technological innovations, artificial intelligence (AI) is rapidly transforming the way that companies interact with their customers. In particular, AI-enabled chatbots are increasingly becoming the leading area of AI use in customer service (Blut et al., 2021). Chatbots provide customers with quick, convenient, and friendly service

through text-based or voice-based chats, which helps reduce the burden of human labor and sometimes replace human employees (Luo et al., 2019). Gartner reported that chatbots will become the principal customer service channel for roughly a quarter of organizations by 2027.<sup>1</sup> Despite the potential benefits of chatbots for customer service, there remains the challenge of ensuring consumers' service evaluation and purchase decisions after chatbot service (Cheng et al., 2022a, b). Some customers feel that chatbots lack the personal touch or empathy that human employees can provide. Moreover, customers with unique or complicated problems are worried chatbots may not have the knowledge or flexibility to provide satisfactory solutions (Fu et al., 2020). Since the chatbot service introduces substantial changes in customers' online shopping experience, it is vital to understand how customers appraise the chatbot service and adapt

---

✉ Bowen Zheng  
bowenzheng@csu.edu.cn

Yang Li  
yangli@hfut.edu.cn

Zhenghua Gan  
gzh@mail.hfut.edu.cn

<sup>1</sup> School of Management, Hefei University of Technology, 193 TunXi Road, Hefei 230009, China

<sup>2</sup> College of Business, Central South University, 932 Lushannan Road, Changsha 410000, China

<sup>1</sup> <https://www.gartner.com/en/newsroom/press-releases/2022-07-27-gartner-predicts-chatbots-will-become-a-primary-customer-service-channel-within-five-years>

to the changes introduced by the intelligent system in their shopping-related decisions.

With its popularity in customer service, chatbots have attracted wide interest from both scholars and practitioners in the e-commerce context (Cheng et al., 2022a, b; Roy & Naidoo, 2021). Prior research has indicated anthropomorphism, which refers to human-like traits exhibited by non-human entities, is a key principle to facilitate customers to interact with chatbots (Lee & Oh, 2021). The human-like traits can evoke a sense of familiarity, comfort, and ease for customers (Cheng et al., 2022a, b). Recognizing the importance of anthropomorphism for chatbots, researchers have started exploring how to maximize the value of chatbots through the design of chatbots' anthropomorphism (Roccapriore & Pollock, 2023), which calls for a further detailed examination of the effect of anthropomorphic traits on customer experience. In practice, companies have also invested in customer service AI primarily to enhance efficiency and foster a sense of intimacy, expecting a detailed guide on the effects of chatbot design in achieving these objectives.<sup>2</sup> As a response to both academic and practical call, the first goal of this study is to investigate how customers evaluate the service of chatbots with anthropomorphism design and how their service evaluation affect their shopping decision process.

Warmth and competence are highlighted as two typical anthropomorphic cues that can deeply engage customers (Cheng et al., 2022a, b; Roy & Naidoo, 2021). Chatbot warmth indicates that customers are perceptive to the chatbots' good intentions to help them, while chatbot competence represents customers' beliefs in chatbots' capacities to act on those good intentions (Roy & Naidoo, 2021). However, equipping chatbot agents with anthropomorphism should be planned and rolled out cautiously because it could be both advantages and threatens to companies. Regarding advantages, when chatbots exhibit human-like capacities such as responsiveness and reliability, customers are more likely to trust the information and assistance provided by the chatbot (Epley et al., 2007), which further enhances their service evaluations. Regarding threats, customers may feel overloaded to process intensive and quick responses from chatbots (Wang et al., 2021). Moreover, machines are commonly believed to lack the mental capability of expressing emotions which is one of the fundamental differences between humans and machines (Luo et al., 2021). Chatbot warmth can evoke a sense of threat to human uniqueness and lead to strong eeriness and aversion toward the machines, which may reduce service evaluations (Han et al., 2022). To comprehensively understand the effect of chatbots on

customers' shopping decisions, the second goal of this study is to empirically validate the dual paths of how chatbot warmth and chatbot competence affect customers' service evaluations through trust in chatbot and chatbot overload.

To provide academical guidance for the research development, the Wixom and Todd model and social information processing theory (SIP) are combinedly applied in this study- the Wixom and Todd model supports the research framework and SIP helps explain how users interpret Anthropomorphism. According to the Wixom and Todd model (Wixom & Todd, 2005), users' behavioral attitudes and intentions are determined by their behavioral beliefs that are driven by their object-based beliefs in service systems, which offers an "object-based beliefs—behavioral beliefs—behavioral attitude—behavioral intention" academic framework. SIP theory further posits that in computer-mediated-communication people engage in a series of cognitive processes (e.g., perceiving, interpreting, and evaluating social information) to process social information to form attitudes and make decisions (Salancik & Pfeffer, 1978; Weisband et al., 1995). According to SIP, customers' attitude and behavior towards chatbot service are results of how they process anthropomorphic cues. In this study, chatbot warmth and chatbot competence are customers' object-based social beliefs about chatbots, which are important social information for customers to process (Wixom & Todd, 2005). Trust in chatbot and chatbot overload are considered customers' behavioral beliefs about chatbots, particularly customers' interpretation of using anthropomorphic service chatbots (Wixom & Todd, 2005). Therefore, the integration of the Wixom & Todd model and SIP provides academic guidance for understanding how social cues of anthropomorphic service systems (i.e., chatbot warmth and chatbot competence) affect users' behavioral beliefs (trust and overload), attitudes (service evaluations), and intentions.

To validate the research design, a pilot study of 33 participants was employed to test the direct effects of chatbot anthropomorphism on customers' service evaluation and purchase intentions. Then a lab-in-the-field experiment with 212 participants and a scenario-based experiment of 124 participants were employed to examine the research model that includes trust in chatbot and chatbot overload as mediating factors. Furthermore, a scenario-based experiment of 232 participants was conducted to uncover the effects of chatbot warmth and chatbot competence on different types of trust in chatbot and chatbot overload. Important findings and implications are discussed.

This study contributes to the extant literature on chatbot anthropomorphism and customer–chatbot interaction, as follows. First, this study reveals that anthropomorphic cues are effective in enhancing customer–chatbot interactions and therefore customers' service evaluation of chatbots in an e-commerce context, which expands current understanding

<sup>2</sup> <https://www.technologyreview.com/2020/04/28/1000675/how-ai-is-changing-the-customer-experience/>

of anthropomorphism and its impact in e-commerce. Second, this study reveals a dual pathway of how chatbot anthropomorphic cues affect customers' service evaluation through trust in chatbot and chatbot overload. It advances a consolidated framework to understand the impacts of chatbot anthropomorphism in the e-commerce context (e.g., Cheng et al., 2022a, b). Third, this study contributes to the contextualisation of the Wixom & Todd model (Wixom & Todd, 2005) and SIP (Salancik & Pfeffer, 1978) in e-commerce chatbots by identifying object-based social beliefs as chatbot anthropomorphism (i.e., chatbot warmth and chatbot competence), behavioral beliefs as trust in chatbot and chatbot overload.

## 2 Literature Review

### 2.1 Chatbot Anthropomorphism

Anthropomorphism is the attribution of human traits (e.g., character, motivation, emotion, and intention) to non-human agents (Epley et al., 2007). Information systems scholars have extensively discussed the effect of chatbot anthropomorphism in online shopping because anthropomorphism is acknowledged to mimic interpersonal interaction and facilitate customer-chatbot interaction (e.g., Schanke et al., 2021; Sharma et al., 2022). A systematic review of prior studies on chatbot anthropomorphism (see Appendix I) indicates that scholars explored various forms of chatbot anthropomorphism which can be classified into human identity cues, verbal cues, and nonverbal cues (Seeger et al., 2021). Some scholars examined chatbot human identity cues such as name (Araujo, 2018; Crolic et al., 2022), gender (Zogaj et al., 2023), human figure icon (Go & Sundar, 2019), and visual appearance cues (Pizzi et al., 2023). Other researchers focused on examining chatbot verbal cues like language style (Araujo, 2018), conversational relevance (Schuetzler et al., 2020), humor (Schanke et al., 2021), and chatbot nonverbal cues such as communication delay (Cheng et al., 2022a, b; Schanke et al., 2021). Moreover, Seeger et al. (2021) posited that the effect of nonverbal cues depends on the co-existence of human identity cues or verbal cues in chatbot communication.

To improve the effectiveness of chatbot anthropomorphism in promoting customer-chatbot interaction, a few studies further explored chatbot warmth and competence as two specific anthropomorphic cues that can shape customers' social judgment when treating chatbots as human-like agents (Cheng et al., 2022a, b; Roy & Naidoo, 2021). For instance, Roy and Naidoo (2021) manipulated warmth and competence as chatbots' human identity cues that can improve customer attitude and purchase intention. Cheng et al. (2022a, b) posited warmth and competence as the

anthropomorphic attributes that can facilitate customer trust. Although warmth and competence cues can be designed into image- and word-based communication to engage customers in social interaction (Roccapriore & Pollock, 2023), few studies have explored the effectiveness of warm and competent verbal or nonverbal cues. Chatbot warmth cues make customers believe that the chatbot has positive intentions for them by displaying a good-natured, tolerant, and enthusiastic image (Cheng et al., 2022a, b). Chatbot competence cues make customers believe that the chatbot can execute positive intentions by showing independence, competitiveness, and confidence (Cheng et al., 2022a, b). This study, therefore, aims to fill out the research gap by manipulating warmth and competence cues in chatbots' nonverbal and verbal communication.

### 2.2 Wixom & Todd Model and Social Information Processing Theory

The Wixom & Todd model posits that users' object-based beliefs influence their behavioral intentions by shaping behavioral beliefs and attitudes about using service systems (Shen et al., 2018; Wixom & Todd, 2005). According to the Wixom & Todd model (Wixom & Todd, 2005), users make behavioral decisions according to their evaluations, experiences, and attitudes toward using the service systems, which are determined by the degree to which users perceive the technological characteristics and functionalities of service systems. Moreover, the Wixom & Todd model (Wixom & Todd, 2005) posits that users' behavioral intentions to use the systems would predict their actual behaviors in the future. Thus, the Wixom & Todd model has been proposed as a useful framework for understanding customers' series reactions to various service systems (e.g., Nguyen et al., 2022; Rhim et al., 2022; Xu et al., 2013).

Complementing the Wixom & Todd model (Wixom & Todd, 2005), SIP provides the theoretical basis for understanding how the social aspects of service systems influence users' attitudes and decisions (Salancik & Pfeffer, 1978). According to SIP, users rely heavily on the social information available to them to shape their perceptions, attitudes, and behaviors (Salancik & Pfeffer, 1978). Users' pursuit of social information helps develop trust relationships (Lu et al., 2019), but also results in information overload problems (Davis & Agrawal, 2018). In this vein, SIP provides the theoretical foundation for explaining how the social cues of chatbots influence customers' interpretation, and evaluation of chatbots and thereby customer purchase decisions.

Prior literature has employed the Wixom & Todd model and SIP to understand customers' responses to service chatbots with different focuses. For instance, Rhim et al. (2022) employed the Wixom & Todd model to evaluate the usability of and user satisfaction with the chatbot systems. Nguyen

et al. (2022) drew upon the Wixom & Todd model to understand customers' attitudes and behaviors toward chatbot services. Hendriks et al. (2020) drew upon SIP to posit that customer-chatbot interaction via written cues can be viewed as a process of building social relationships. Schanke et al. (2021) employed SIP to propose that anthropomorphism can be viewed as the social cues of chatbots to make social interpretations in computer-mediated communication. The above studies suggest that the question of how chatbot anthropomorphism affects customers' behavioral attitudes and intentions remains unclear in extant literature. This study draws upon the Wixom & Todd model and SIP to identify the object-based social beliefs, behavioral beliefs, and behavioral attitudes that influence customers' behavioral intentions towards chatbots in online service.

### 2.2.1 Chatbot Warmth and Competence as Object-Based Social Beliefs

Object-based social beliefs refer to users' perceptions of the social aspects of service systems, such as perceived information quality (Wixom & Todd, 2005). According to the Wixom & Todd model (Wixom & Todd, 2005), customers first observe the technological characteristics and functionalities of systems and become aware of object-based social beliefs in their minds. Prior research has contextualised object-based social beliefs about chatbot services. For instance, Nguyen et al. (2022) theorised that the information and system quality of chatbot applications are object-based social beliefs because they describe customers' perceptions of chatbot systems' capabilities to convey information or services. Walter et al. (2022) theorised that lay beliefs about artificial intelligence expert systems are object-based social beliefs because they describe customers' beliefs in artificial intelligence algorithms that make an optimal choice. Such object-based social beliefs influence customers' advocacy of service providers (Nguyen et al., 2022) and the adoption of algorithmic advice (Walter et al., 2022) in the context of chatbot services.

These previous studies indicated that object-based social beliefs about chatbot services play a significant role in explaining customer behavior. However, little research has identified specific object-based social beliefs that explain customer behavior toward online service chatbots. Based on the SIP (Salancik & Pfeffer, 1978), we theorised that perceived chatbot anthropomorphic traits are object-based social beliefs because the anthropomorphic traits of chatbots convey the social information of e-commerce chatbots. Anthropomorphic chatbot traits refer to the human-like traits (e.g., warmth- and competence-based communication cues) that induce customers to perceive non-human agents (e.g., chatbots) as human agents (Cheng et al., 2022a, b). Specifically, chatbot warmth leads customers to believe that their

intentions are positive like humans (Chien et al., 2022), such as offering good-natured or enthusiastic service. Chatbot competence can be viewed as customers' beliefs that chatbots have a strong capacity to act on positive intentions as a human (Ahmad et al., 2022), such as showing confidence and competitiveness. Thus, chatbot warmth and chatbot competence are identified as the object-based social beliefs of chatbots in online services.

### 2.2.2 Trust in Chatbot and Chatbot Overload as Behavioral Beliefs

Behavioral beliefs are viewed as users' evaluations or experiences of using the system, such as perceived usefulness (Wixom & Todd, 2005). Further, the Wixom & Todd model (Wixom & Todd, 2005) posits that object-based social beliefs influence users' behavioral attitudes by promoting behavioral beliefs. Despite the significance of behavioral beliefs in predicting behavioral attitudes, the specific type of behavioral beliefs that can explain customers' behavioral attitudes toward chatbot service systems remains ambiguous in the literature. According to the literature on SIP, trust can be theorised as a behavioral belief pursued by customers because trust implies a close relationship in computer-mediated communication (Jarvenpaa & Leidner, 1999). Trust is proposed to influence customers' behavioral attitudes toward the voice user interface on smartphones (Nguyen et al., 2019) and their intentions to adopt personal intelligent agents (Moussawi et al., 2021). In the chatbot service context, Behera et al. (2021) examined the influence of trust in chatbot on customers' attitudes toward chatbots. Moreover, trust in chatbot indicates the evaluative aspect of behavioral beliefs because it reflects customers' positive expectations of the outcome of using chatbots (Kyung & Kwon, 2022). Based on customers' evaluations of different aspects of the chatbot, trust in this technology may include an objective assessment of competence (i.e., cognitive trust) and a subjective assessment of the emotional connection (i.e., affective trust) to it (Hildebrand & Bergner, 2021).

On the other hand, the term chatbot overload refers to customers' negative experiences in processing excessive information beyond their mental resources or personal expertise (Wang et al., 2021). Therefore, chatbot overload is construed as an experiential aspect of behavioral beliefs because it reflects customers' negative experiences of using chatbot technology. Prior research has identified information overload in human-chatbot interaction (Luo et al., 2021) and examined the role of technology overload in the effect of chatbot technology utilization on service outcomes (Kim et al., 2022). Literature on SIP also highlights technology overload as a persistent problem of information processing because users heavily rely on all available social information to make further decisions

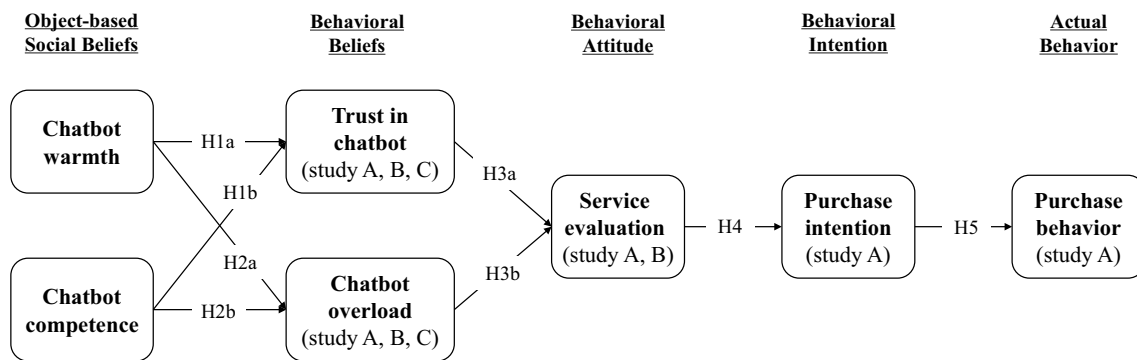


Fig. 1 Research model

(Davis & Agrawal, 2018). As a negative reaction to chatbot technology, chatbot overload may include cognitive symptoms of errors in decisions (i.e., cognitive overload) and emotional symptoms, such as stress and frustration (i.e., emotional overload) (Rutkowski et al., 2013; Saunders et al., 2017). The literature on the Wixom & Todd model posits that behavioral beliefs can be understood as users' evaluations of the outcomes or experience of using technology (Shen et al., 2018). Thus, this study considered trust in chatbot and chatbot overload as evaluative and experiential aspects of customer behavioral beliefs.

### 2.2.3 Service Evaluation as Behavioral Attitude

Behavioral attitudes represent the users' attitudes toward using the system, such as user satisfaction (Wixom & Todd, 2005). Literature on the Wixom & Todd model argues that behavioral beliefs affect behavioral intention by shaping behavioral attitudes (Xu et al., 2013). The significance of customer attitude in predicting customer behavior calls for the consideration of service evaluation of chatbot technology (Behera et al., 2021; Han et al., 2022). A favorable evaluation indicates customers' positive attitudes toward experienced service (Keh & Sun, 2018). In chatbot service, Han et al. (2022) found that customers' positive emotions aroused by chatbot services lead to improved service evaluation. Roy and Naidoo (2021) reported that chatbot anthropomorphism improves customer attitudes toward the brand. Behera et al. (2021) indicated that attitude toward chatbot technology has a positive influence on customer adoption of chatbots. However, the role of behavioral attitude between behavioral beliefs and behavioral intentions has not been examined in the context of chatbot service. Therefore, this study viewed service evaluation as the behavioral attitude towards chatbot service that explains how customers' behavioral beliefs induce behavioral intentions.

## 3 Hypothesis Development

In summary, the integration of the Wixom & Todd model and SIP provides a comprehensive framework for theorising our research model. Based on this model, we first examine how object-based social beliefs (e.g., chatbot warmth, chatbot competence) influence behavioral beliefs (e.g., trust in chatbot and chatbot overload) and thereby behavioral attitude (e.g., service evaluation). Moreover, we examined the effect of behavioral attitude on behavioral intention (e.g., purchase intention) and then actual behavior (purchase behavior). Figure 1 shows the research model.

### 3.1 Anthropomorphism and Trust in Chatbot

Literature on SIP highlights anthropomorphic traits as the social cues that can develop close relationships in customer-chatbot communication (Schanke et al., 2021). Given machines are commonly believed to lack the mental capability of expressing emotions (Luo et al., 2021), warmth traits, such as empathetic and personable words, are applied in chatbot scripts to convey kindness or good faith in order to please customers. Warm traits lead customers to perceive chatbots as human-like agents because they mimic customers' social interactions with familiar people, such as friends and family members. When chatbots respond with communication modes characterised by warm traits, customers' cognitive needs for product-related information and emotional needs for caring could be satisfied (Gong et al., 2016). As a result, chatbot warmth induces customers to feel that chatbots are trustworthy agents.

Moreover, chatbot warmth has the potential to encourage social proximity between customers and chatbots during service-delivery conversations (Ehrke et al., 2020). Customers then perceive chatbots in the same way as they perceive human service agents, which builds customer trust in chatbot. In contrast, when customers perceive less warmth in conversations with chatbots, they feel that chatbots are cold

machines. It would then be difficult for customers to engage in social interactions with chatbots. Consequently, customers are less likely to build trust in chatbot. The literature on the Wixom & Todd model posits that trust reflects users' behavioral beliefs in chatbots' good intentions because of the latter's anthropomorphic traits (Cheng et al., 2022a, b). Chatbot warmth as customers' object-based social beliefs about chatbots could elicit positive behavioral beliefs about them, which is considered customers' trust in this study. Therefore, the following hypothesis was proposed:

*H1a. Chatbot warmth is positively related to trust in chatbot.*

Trust in chatbot could also be established by customers' positive expectations of chatbot competence, which has been posited as an important anthropomorphic trait (Cheng et al., 2022a, b). Literature on SIP viewed competence traits as a form of social information that influence people's social information processing in an interpersonal relationship (Chan & Ybarra, 2002). Chatbot competence captures customers' perceptions of chatbots' capacity to deliver skilled services. Chatbots with high levels of competence are viewed as confident and competitive service agents that can diagnose customers' questions and provide information that is useful in maintaining high-quality conversations. By creating positive expectations in the customer's mind, chatbot competence could lead to increased trust in chatbot. In contrast, low competence, such as failing to understand customers' questions and providing inaccurate responses, might induce customers to feel that chatbots have less ability and efficiency in task solving. Therefore, unsatisfactory performance caused by the incompetence of chatbots could damage customers' trust in this technology (Jiang, 2023). Therefore, the following hypothesis was proposed:

*H1b. Chatbot competence is positively related to trust in chatbot.*

### 3.2 Anthropomorphism and Chatbot Overload

In e-commerce services, chatbots often respond to customers in text-based communications. Chatbot overload occurs when users feel overwhelmed or burdened by excessive interactions with the chatbot (Saunders et al., 2017). For example, the chatbot provides responses that are too lengthy or complex, or fails to understand and address the user's needs effectively. Large amounts of information require high levels of cognitive capabilities, which often exceed the mental efforts that customers are willing to make in shopping context (Fu et al., 2020). Compared with robotic language, conversations imbued with warm traits resemble the natural communication that customers usually have in daily life and

can easily accommodate. The familiarity leads customers to automatically filter out relevant information that may help them in online service inquiries. According to SIP theory, when a chatbot exhibits warmth, such as using friendly language and showing empathy, users would perceive these social cues as positive emotional connection which would help reduce the likelihood of feeling overloaded by the interactions with the chatbot (Davis & Agrawal, 2018). Therefore, the anthropomorphic traits of chatbots are expected to decrease customers' feelings of overload when they communicate with chatbots. That is, chatbot warmth could increase customers' readiness and familiarity with engaging in conversations with chatbots, thereby relieving the pressure on customers to process textual communication. Therefore, the following hypothesis was proposed:

*H2a. Chatbot warmth is negatively related to chatbot overload.*

When communicating with chatbots in e-commerce transactions, customers expect chatbots to be competent in understanding their needs and responding accurately. Customers would feel reflex and easy when the chatbots are professional and mature agents that can meet their service requests (Nguyen et al., 2022). That is, chatbot competence would induce customers to consider that the information and services delivered by chatbots are necessary to facilitate online shopping, which could reduce customers' sense of chatbot overload. In contrast, a low level of competence would make customers feel overloaded because they have to process inaccurate and irrelevant information when the chatbot does not understand and respond to them correctly. Moreover, the low competence of chatbots would force customers to seek employee services, thereby increasing customer overload. Therefore, the following hypothesis was proposed:

*H2b. Chatbot competence is negatively related to chatbot overload.*

### 3.3 Trust in Chatbot, Chatbot Overload, and Customer Behavior

As an emerging technology, chatbots have the potential for positive performances, including trust relationships (Mostafa & Kasamani, 2022), but they also have negative aspects of emerging technologies, such as technology overload (Kumar et al., 2022; Sun et al., 2022). According to the Wixom & Todd model, customers' behavioral intentions regarding chatbot usage are determined by their behavioral beliefs, which are based on their evaluations of the outcomes or experiences of using chatbots (Wixom & Todd, 2005). Trust in chatbot (Mostafa & Kasamani, 2022) and chatbot overload (Sun et al., 2022) are two crucial

behavioral beliefs about chatbot usage that could influence customers' intentions to use chatbots.

Chatbots are embedded in online stores as a customer self-service technology that replaces human employees (Li et al., 2021). In online shopping, when customers trust chatbots, they have an objective assessment of the competence and quality of the chatbot in interaction (i.e., cognitive trust) and a subjective evaluation of the emotional bond with the chatbot (emotional trust) (Hildebrand & Bergner, 2021). In other words, trust in a chatbot can be conceptualised as customers' willingness to be vulnerable to the actions of the chatbot because they have positive expectations of this technology (Kim et al., 2009). As a result, customers seamlessly transit to a chatbot agent service. Prior research has posited that trust in chatbot reflects the customer's belief that the chatbot is reliable and understandable (Li et al., 2021), which facilitates the customer's usage intention (Mostafa & Kasamani, 2022). Trust in chatbot implies that customers have confidence in the chatbot's competence to provide accurate and safe service (Chen et al., 2020). Based on positive expectations of acquiring high-quality service and establishing a close relationship, trust in chatbot should lead to an increased service evaluation. Therefore, the following hypothesis was proposed:

*H3a. Trust in chatbot is positively related to service evaluation.*

Technology overload has often been conceptualised as the various cognitive and emotional symptoms that result from interacting with information and communication technologies (Saunders et al., 2017). As an emerging self-service technology in online shopping, chatbots have been criticised because of wake-up errors, insensitive responses, and irrelevant answers, all of which impose cognitive and emotional overloads on customers (Sun et al., 2022). When customers interact with chatbots, cognitive overload is manifested by information that exceeds the customer's mental resources or personal expertise (Fu et al., 2020), whereas emotional overload implies customers' frustration, pressure, and confusion (Rutkowski et al., 2013), both of which are related to negative outcomes. Prior research has found that chatbot overload negatively affects customers' satisfaction and reduces customers' continued intention to use chatbots (Sun et al., 2022). Because of their negative beliefs about chatbots, such as feeling overloaded by using chatbot technology, customers are inclined to have a low level of service evaluation of chatbot service. Therefore, the following hypothesis was proposed:

*H3b. Chatbot overload is negatively related to service evaluation.*

### 3.4 Effect of Service Evaluation on Purchase Intention

When customers have a positive service evaluation of chatbots, it indicates that the chatbot provided a satisfactory and efficient interaction. When customers have their problems resolved promptly and effectively, it creates a positive impression and boosts their confidence in the store. This enhanced customer experience can lead to positive perceptions of the brand and the online store, increasing their confidence in making a purchase (Zarei et al., 2019). As a result, a positive service evaluation would increase the likelihood of purchase intention. On the contrary, if a customer has a negative experience with a chatbot, such as receiving inaccurate information or experiencing repeated technical glitches, it can undermine their confidence in the purchase process, leading to decreased purchase intention. Previous studies also suggested the positive effect of service evaluation on customers' purchase intention (Behera et al., 2021; Han et al., 2022). Therefore, this study proposed that:

*H4. Service evaluation is positively related to purchase intention.*

### 3.5 Effect of Purchase Intention on Purchase Behavior

When customers have a strong purchase intention in the e-commerce context, it indicates a certain level of commitment to making the purchase (Morrison, 1979). This commitment translates into a higher likelihood of following through with the intended purchase and converting the intention into actual purchase behavior. Customers have confidence in the expected satisfaction and value they will derive from the purchase (Zarei et al., 2019). When customers believe that the product or service will meet their needs and provide a positive experience, it increases their intention to purchase and subsequently, their actual purchase behavior (Newberry et al., 2003). Therefore, this study proposed that:

*H5. Purchase intention is positively related to purchase behavior.*

## 4 Research Design

The unit of analysis of this study is a customer with experience in using chatbots in online shopping. An online lab-in-the-field chatbot experiment (study A) and two scenario-based chatbot experiments (study B, C) were conducted to examine the research model. Appendix II presents the research design including three quantitative studies. Study A uses a lab-in-the-field experiment with a student

sample which allows for observing and testing participants' actual behavior (i.e., adding a shopping cart) in actual online stores. Study B uses a scenario-based experiment with the non-student sample which helps further test and generalize the differential effects of four versions of chatbots. Study C uses a scenario-based experiment with the non-student sample which helps further uncover the differential effects of five versions of chatbots. The experiential tasks were designed as purchasing experiential products (i.e., t-shirts, shampoo) which are considered appropriate to test the effectiveness of chatbots in online services (Zhu et al., 2022).

By imitating DianXiaoMi in Taobao,<sup>3</sup> one of the biggest e-commerce platforms in China, four versions of chatbots characterised by different anthropomorphic cues (i.e., none, warmth, competence, and a hybrid of warmth and competence) were developed. To explain the relationship between the three studies and how they serve the research objectives, we have added a table for clarification. As shown in Appendix III, paralanguage cues (e.g., emoticons, number of words, tone) were used to describe different versions of chatbots under experimental conditions, which were then designed to simulate real-life shopping scenarios.

To measure chatbot warmth and chatbot competence in study A, B, and C, a three-item scale and a three-item scale were adapted from Fiske et al. (2002). We conceptualised trust in chatbot and chatbot overload as second-order formative constructs in study A, B and as first-order reflective constructs in study B to examine whether measurement scale could influence the results. Service evaluation in study A, B was evaluated by one item adopted from Keh and Sun (2018). In addition, study A measured purchase intention with one item and purchase behavior with a dummy variable using a self-developed scale. All principal constructs (except for purchase intention and purchase behavior) were measured by a seven-point Likert scale ranging from 1 ("strongly disagree") to 7 ("strongly agree"). Following the reflective–formative construct operationalisation guidelines in the information system discipline (Cenfetelli & Bassellier, 2009; Petter et al., 2007), we conceptualised all first-order principle constructs as reflective constructs. Appendix IV shows the scales of focal constructs for testing the hypotheses and the experimental conditions (i.e., perceived informativeness, perceived realism, perceived difficulty).

We included gender, age, education, prior online shopping experience, and prior chatbot experience as the control variables, based on prior studies on customers' reactions to chatbots (Roy & Naidoo, 2021). Gender was measured by one item: 0 = female and 1 = male. Age was measured by one item: 1 = 18–24, 2 = 25–30, 3 = 31 and above. Education

was measured by one item: 1 = Undergraduate and below, 2 = Graduate and above. The prior online shopping experience was measured using one item: "The number of years I have used online shopping" (from "1 = <2" to "5 = >8"). Prior chatbot experience was measured using one item: "The number of years I have used chatbot" (from "1 = <2" to "5 = >8"), which was adapted from Shen et al. (2018).

We conducted a subject matter expert review to ensure the accuracy of all scales. The questionnaire items were translated into Chinese and then back-translated into English to ensure equivalent meanings across the two versions. Before the actual data collection, two experienced researchers from the information systems and marketing disciplines were invited to check the expression and clarity of the questionnaire.

To calculate an appropriate number of research samples for testing the research model, we conducted an a priori statistical power analysis. The maximum number of predictors in the research model was 12. Assuming a medium effect size ( $f^2 = 0.30$ ), a minimum sample size of 111 respondents to enable an alpha level of 0.05 and a power of 0.95 were required for the research model (Cohen, 1988).

## 4.1 Study A

### 4.1.1 Pilot Study

The pilot study aimed to test the lab–in–the–field experiment of study A and investigate the premise that whether chatbots demonstrating higher chatbot warmth and chatbot competence can elicit higher service evaluation and customer purchase. However, the mediating constructs (i.e., trust in chatbot and chatbot overload) were not included (Schuetzler et al., 2020). The procedures applied in study A included three stages. In Stage 1, we recruited participants from universities in Mainland China. To be eligible for the experiment, the participants were required to answer an enrolment questionnaire that included questions related to demographic information, personal contact, and times available for participating in the experiment. In Stage 2, we recruited four research assistants, including one of the authors, to manage the four experimental conditions. The research assistants were trained to contact and instruct the participants and record the experimental details using Tencent Conference.<sup>4</sup> With the aid of the research assistants, the participants were randomly allocated to one of the Taobao stores. They then communicated with chatbots in a real shopping environment. In Stage 3, we used texts and pictures to introduce the experimental procedure and shopping task.

<sup>3</sup> Taobao (<https://www.taobao.com/>) features designing experimental conditions, i.e., manipulating chatbots cues.

<sup>4</sup> Tencent Meeting (<https://meeting.tencent.com/>) features communicating screen sharing and screen recording.



Before the experiment, the participants were informed of the anonymous, voluntary, authentic, non-commercial nature of the experiment. Regarding the shopping tasks, the participants were asked to purchase uniforms (i.e., t-shirts) for their classes, and the chatbots in the online stores were prepared for their inquiries. To increase the response rate, the participants were offered an incentive of 30~50 RMB. To ensure that the participants engaged fully in the experiment, they were asked to check at least five of the 15 questions about the target shopping task. They then used these questions to converse with the chatbots. After the conversations with the chatbots, a post-experiment questionnaire that includes all research constructs was answered by the participants. Participants were also asked to answer questions related to perceived informativeness, perceived difficulty, perceived realism of experimental conditions, and attention check questions.

A data sample of 33 participants was collected from universities in Mainland China. We conducted an exploratory analysis and a hierarchical regression analysis to evaluate the psychometric properties and co-relationships among chatbot warmth, chatbot competence, service evaluation, and purchase intention. Initially, we performed Kaiser–Olkin (KMO) and Bartlett’s tests using SPSS 26 (George & Mallery, 2019) to conduct a descriptive factor analysis (Lewis et al., 2005). The result of the KMO test was 0.81, and the result of Bartlett’s test of sphericity was significant at  $p=0.00$ , thus validating the adequacy of the data sample for exploratory factor analysis. We validated the scales of the nine items for internal reliability, convergent validity, and discriminant validity (Lewis et al., 2005). As shown in Table VI (Appendix V), the average variance extracted (AVE) values were higher than 0.50, the composite reliability (CR) values were 0.92, and the Cronbach’s alpha values were 0.86, which was higher than the threshold value of 0.70 (Nunnally, 1978), indicating satisfactory internal validity. The factorial loadings ranged from 0.82 to 0.93, indicating acceptable convergent validity. The correlations among chatbot warmth, chatbot competence, service evaluation, and purchase intention ranged from 0.33 to 0.76, and the square roots of AVEs were higher than the intra-construct correlations, indicating satisfactory discriminant validity (Hair et al., 2006).

We examined the direct effects of chatbot warmth and chatbot competence on service evaluation and purchase intention by applying hierarchical regression analysis using SPSS 26. Two models were estimated. As shown in Table V2 (Appendix V), when chatbot warmth and chatbot competence were included to explain service evaluation, the  $R^2$  value increased by 0.44 at a significant level ( $F=10.10$ ,  $p<0.001$ ), chatbot warmth was nonsignificant ( $\beta=0.11$ ,  $p>0.05$ ) while chatbot competence was significantly ( $\beta=0.61$ ,  $p<0.001$ ) related to service evaluation. When

chatbot warmth and chatbot competence were included to explain purchase intention, the  $R^2$  value increased by 0.16 at a significant level ( $F=2.92$ ,  $p<0.05$ ), chatbot warmth was nonsignificant ( $\beta=-0.23$ ,  $p>0.05$ ) while chatbot competence was significantly ( $\beta=0.75$ ,  $p<0.05$ ) related to purchase intention. Thus, we concluded that chatbot competence and chatbot warmth influence service evaluation and purchase intention through other plausible mechanisms.

#### 4.1.2 Formal Study

Following the modified procedure based on the pilot study, study A was conducted from April 1st to April 26th, 2022. Data were collected through an embedded questionnaire conducted on Wenjuan.com, which features delivering experimental conditions, i.e., Taobao store URL, remotely. A data sample of 226 participants was collected from universities in Mainland China. Several criteria were employed to fill out the qualified responses. Among the 226 participants, five failed to receive responses or received delayed responses from chatbots because of technical problems, three participants did not ask the chatbot questions, five participants did not pass the attention check, and one participant gave regular answers to the research constructs in this study. Thus, 212 valid responses ( $N_{\text{control}}=53$ ,  $N_{\text{warmth}}=54$ ,  $N_{\text{competence}}=55$ ,  $N_{\text{hybrid}}=50$ ) were retained for further data analysis. To determine whether 212 responses were sufficient for the data analysis, we used the G\*Power 3.1.9.7 software to implement a post hoc statistical power analysis (Faul et al., 2007). The average effect size of the relationships presented in this research model was 0.04, with an alpha level of 0.05 and a power of 0.99, which was higher than the threshold value of 0.80. Thus, our sample size was sufficient for the data analysis.

Table A1 presents the demographic profiles of the respondents. Among the 212 respondents, 72.60% were female, 78.80% were aged above 18 years and below 24 years, 50.50% had a bachelor’s degree or below, 94.80% had more than two years of prior online shopping experience, 71.20% had more than two years of prior chatbot experience.

Following prior research on testing the successful randomization of experimental conditions and the effects of control variables (Rzepka et al., 2022), we used analyses of variance (ANOVAs) for the metric variables and Fisher’s exact tests for the categorical variables. First, the levels of chatbot warmth ( $F(3, 208)=20.57$ ,  $p=0.00$ ) and chatbot competence ( $F(3, 208)=6.96$ ,  $p=0.00$ ) were found to differ significantly between the control and experimental groups. Second, the homogeneity of the conversation content was ensured because the participants reported only a minimal difference in perceived informativeness between the control and hybrid groups ( $F(3, 208)=2.80$ ,  $p=0.04$ ). Moreover, the participants reported that they felt the

experimental conditions were realistic and easy to understand, and there was no significant difference across the four experimental conditions regarding perceived realism ( $F(3, 208) = 0.324, p = 0.81$ ) and perceived difficulty ( $F(3, 208) = 0.82, p = 0.48$ ). Third, the results showed that there was no significant group difference in gender ( $p = 0.99$ ), age ( $p = 0.50$ ), education ( $p = 0.32$ ), prior online shopping experience ( $p = 0.94$ ), prior chatbot experience ( $p = 0.23$ ). Thus, we concluded that the treatment was successful and the control variables would not confound the experimental effects.

Following prior research, we simultaneously employed PLS-SEM to test the mediating effects of trust in chatbot and chatbot overload (Schuetzler et al., 2020). In the information system research community, PLS-SEM has been widely used with complex models that include multiple mediators and dependent variables. Moreover, PLS-SEM efficiently models relatively small samples with stable statistical quality (Reinartz et al., 2009). We applied the Smart-PLS 3.27 software package to analyses 5,000 bootstrapping resamples. Following Hair et al.'s (2021) two-step analytical procedure, we determined the results of the measurement model and structural model. We validated the measurement model of first-order reflective constructs in terms of internal reliability, as follows: Cronbach's alpha  $> 0.70$ ; composite reliability (CR)  $> 0.70$ ; average variance extracted (AVE)  $> 0.50$ ; convergent validity, loadings  $> 0.60$ ; and discriminant validity, square roots of the AVEs  $>$  construct correlations (Fornell & Larcker, 1981). As shown in Table A2, Cronbach's alpha ranged from 0.83–0.85, CR ranged from 0.90–0.91, and AVE ranged from 0.75–0.77, indicating satisfactory internal reliability. The item loadings ranged from 0.83–0.91, which indicated acceptable convergent validity. We validated the measurement model of the second-order formative constructs in terms of weight size and weight significance (Petter et al., 2007). The size and significance of the formative indicators' weights revealed their respective roles in significantly determining the formation of a formative construct. As shown in Table A3, all formative indicators had statistically significant weights, with acceptable weight sizes.

To determine whether multicollinearity existed in this study, we examined the variance inflation (VIF) factor and the tolerance value of all constructs. As suggested by Mason and Perreault (1991), multicollinearity is less likely to be a serious problem if the VIF value is lower than 10.00 for reflective indicators or lower than 3.33 for formative indicators. The tolerance value of both the reflective and formative indicators should be higher than 0.10. As shown in Table A2, the VIF of all reflective constructs ranged from 1.93–2.79, whereas the tolerance value ranged from 0.36–0.52. Similarly, the VIF of all formative constructs ranged from 1.71–2.29, whereas the tolerance value ranged from 0.44–0.58. Therefore, multicollinearity was not a concern in this study.

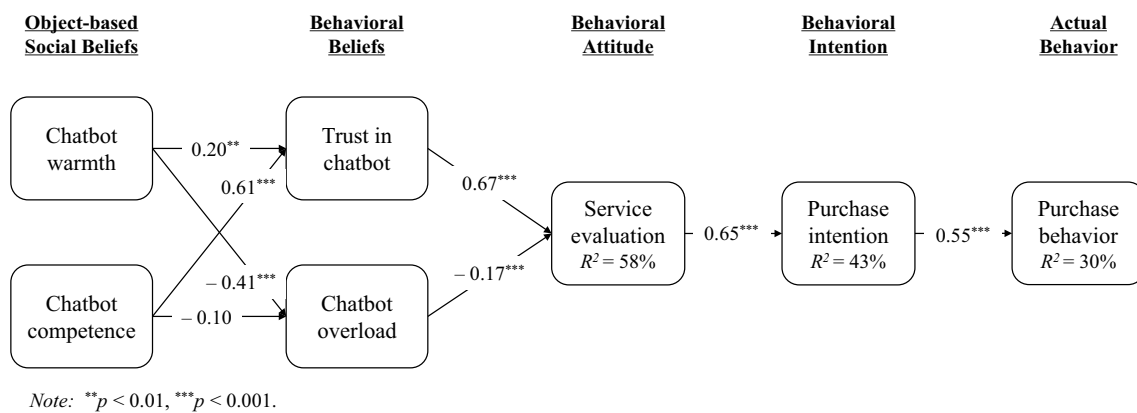
Moreover, discriminant validity was established because the square roots of the AVEs for all constructs were greater than the construct correlations shown in Table A4. Combined, these results indicate that the reliability and validity of the measurement model were satisfactory.

To address potential concerns about using perceptual measures and self-reported data, we conducted analyses to determine nonresponse bias and common method bias (see Appendix VI). The results showed that these biases were not serious concerns in this study.

Because the research model included second-order constructs, we followed prior studies and applied the indicator reuse approach by computing PLS algorithms to calculate the latent variable score of the focal constructs, which we used to test the main effects using PLS-SEM with 5,000 bootstrapping resamples (Lowry & Gaskin, 2014).

**Main Effects** Consistent with H1a and H1b, chatbot warmth (H1a:  $\beta = 0.20, p < 0.01$ ) and chatbot competence (H1b:  $\beta = 0.61, p < 0.001$ ) were positively related to trust in chatbot. Consistent with H2a, chatbot warmth (H2a:  $\beta = -0.41, p < 0.001$ ) was negatively related to chatbot overload. However, the relationship between chatbot competence and chatbot overload (H2b:  $\beta = -0.10, p > 0.05$ ) was nonsignificant; thus, H2b was rejected. Consistent with H3a and H3b, trust in chatbot was positively associated with service evaluation (H3a:  $\beta = 0.67, p < 0.001$ ), whereas chatbot overload was negatively related to service evaluation (H3b:  $\beta = -0.17, p < 0.001$ ). Moreover, service evaluation was positively related to purchase intention (H4:  $\beta = 0.65, p < 0.001$ ), while purchase intention was positively related to purchase behavior (H5:  $\beta = 0.55, p < 0.001$ ), thus, H4 and H5 were supported. The model explained 58%, 43%, and 30% of the variance in service evaluation, purchase intention, and purchase behavior, respectively. Figure 2 illustrates the results of the direct effects of study A.

**Mediation Effects** To examine the mediating roles of trust in chatbot and chatbot overload, we used the latent variable score derived in SmartPLS to conduct a mediating analysis using SPSS. Specifically, we used PROCESS Model 4 for SPSS to obtain the regression coefficients (Hayes & Scharkow, 2013). The standard two-tailed test was conducted using the standard SPSS macro with 5,000 bootstrapping resamples and bias-corrected confidence estimates [95% CI]. As shown in Table A5, the results indicated that the indirect effects of chatbot warmth on service evaluation through trust in chatbot ( $\beta = 0.15, CI_{95\%} = [0.0264, 0.1655], p < 0.05$ ) and chatbot overload ( $\beta = 0.06, CI_{95\%} = [0.0151, 0.1088], p < 0.05$ ) were significant, but the direct effects were nonsignificant ( $\beta = 0.12, CI_{95\%} = [-0.0008, 0.2362], p > 0.05$ ). This result indicated that trust in chatbot and chatbot overload fully mediated the influence of chatbot



**Fig. 2** Structural model of study A

warmth on service evaluation. Similarly, the indirect effects of chatbot competence on service evaluation through trust in chatbot ( $\beta = 0.28$ ,  $CI_{95\%} = [0.1717, 0.3811]$ ,  $p < 0.05$ ) were significant, while the direct effects were significant ( $\beta = 0.21$ ,  $CI_{95\%} = [0.0732, 0.3451]$ ,  $p < 0.05$ ). This result indicated that trust in chatbot partially mediated the influence of chatbot competence on service evaluation.

## 4.2 Study B

Following the procedure in study A, study B was conducted from May 1st to May 5th, 2023. Data were collected through a scenario-based questionnaire conducted on wjx.cn, which includes four customer-chatbot dialogue scenarios (see Appendix III). A data sample of 191 participants was collected in Mainland China. After filling out the qualified responses using attention check questions (e.g., *whether the chatbot used emoticons in dialogue?*), 124 ( $N_{\text{control}} = 30$ ,  $N_{\text{warmth}} = 35$ ,  $N_{\text{competence}} = 33$ ,  $N_{\text{hybrid}} = 26$ ) valid sample was used to further data analysis. To determine whether 124 responses were sufficient for the data analysis, we used the G\*Power 3.1.9.7 software to implement a post hoc statistical power analysis (Faul et al., 2007). The average effect size of the relationships presented in this research model was 0.05, with an alpha level of 0.05 and a power of 0.97, which was higher than the threshold value of 0.80. Thus, our sample size was sufficient for the data analysis. Table B1 presents the demographic profiles of the respondents. Among the 124 respondents, 66.90% were female, 90.30% were aged above 18 years and below 24 years, 70.20% had a bachelor's degree or below, 98.40% had more than two years of prior online shopping experience, 58.90% had more than two years of prior chatbot experience.

Similar to Study A, we used analyses of variance (ANOVAs) for the metric variables and Fisher's exact tests for the categorical variables. First, the levels of chatbot warmth ( $F(3, 120) = 14.98$ ,  $p = 0.00$ ) and chatbot competence ( $F(3, 120) = 5.08$ ,  $p = 0.00$ ) were found to differ significantly between the control and experimental groups. Second, the homogeneity of the conversation content was ensured because the participants reported only a minimal difference in perceived informativeness between the control and competence groups ( $F(3, 120) = 4.01$ ,  $p = 0.01$ ). Moreover, the participants reported that they felt the experimental conditions were realistic and easy to understand, and there was no significant difference across the four experimental conditions regarding perceived realism ( $F(3, 208) = 0.77$ ,  $p = 0.51$ ). Third, the results showed that there was no significant group difference in gender ( $p = 0.51$ ), age ( $p = 0.82$ ), education ( $p = 0.74$ ), prior online shopping experience ( $p = 0.63$ ), prior chatbot experience ( $p = 0.92$ ). Thus, we concluded that the treatment was successful and the control variables would not confound the experimental effects.

As shown in Table B2, Cronbach's alpha ranged from 0.78–0.90, CR ranged from 0.87–0.94, and AVE ranged from 0.70–0.84, indicating satisfactory internal reliability. The item loadings ranged from 0.71–0.96, which indicated acceptable convergent validity. Moreover, Table A2 also shows that the VIF of all reflective constructs ranged from 1.21–1.50, whereas the tolerance value ranged from 0.67–0.83. Therefore, multicollinearity was not a concern in this study. Moreover, discriminant validity was established because the square roots of the AVEs for all constructs were greater than the construct correlations shown in Table B3. Combined, these results indicate that the reliability and validity of the measurement model were satisfactory. The results of nonresponse bias and common method bias (see Appendix VI) showed that these biases were not serious in Study B.

**Main Effects** Consistent with H1a and H1b, chatbot warmth (H1a:  $\beta = 0.31$ ,  $p < 0.05$ ) and chatbot competence (H1b:  $\beta = 0.32$ ,  $p < 0.01$ ) were positively related to trust in chatbot. However, chatbot warmth (H2a:  $\beta = -0.14$ ,  $p > 0.05$ ) was

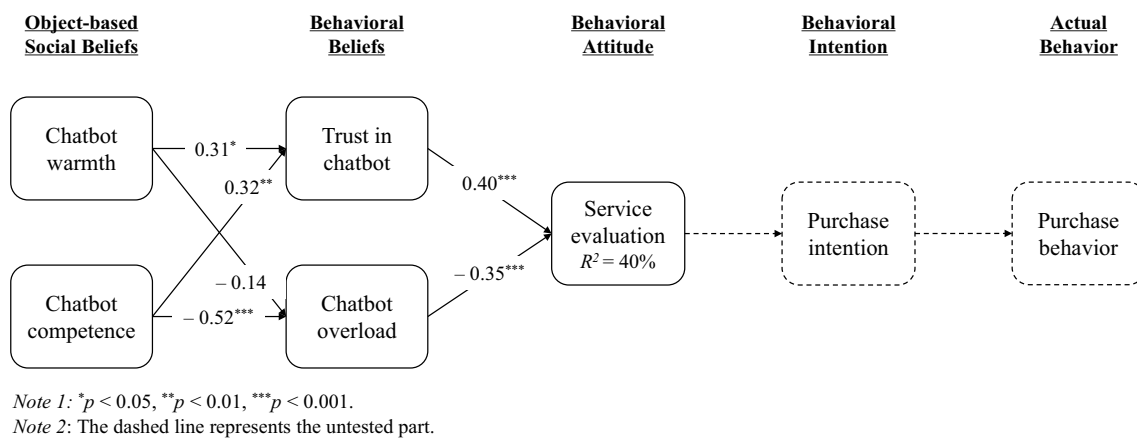


Fig. 3 Structural model of study B

non-significantly related to chatbot overload; thus, H2a was rejected. The relationship between chatbot competence and chatbot overload (H2b:  $\beta = -0.52$ ,  $p < 0.001$ ) was significant; thus, H2b was supported. Consistent with H3a and H3b, trust in chatbot was positively associated with service evaluation (H3a:  $\beta = 0.41$ ,  $p < 0.001$ ), whereas chatbot overload was negatively related to service evaluation (H3b:  $\beta = -0.35$ ,  $p < 0.001$ ). The model explained 40% of the variance in service evaluation. Figure 3 illustrates the results of the direct effects of Study B.

**Mediation Effects** As shown in Table B4, the results indicated that the indirect effects of chatbot warmth on service evaluation through trust in chatbot ( $\beta = 0.08$ ,  $CI_{95\%} = [0.0006, 0.2099]$ ), but the direct effects were nonsignificant ( $\beta = 0.13$ ,  $CI_{95\%} = [-0.0431, 0.2963]$ ,  $p > 0.05$ ). This result indicated that trust in chatbot fully mediated the influence of chatbot warmth on service evaluation. Similarly, the indirect effects of chatbot competence on service evaluation through trust in chatbot ( $\beta = 0.08$ ,  $CI_{95\%} = [0.0078, 0.2141]$ ,  $p < 0.05$ ) and chatbot overload ( $\beta = 0.09$ ,  $CI_{95\%} = [0.0042, 0.2296]$ ,  $p < 0.05$ ) were significant, while the direct effects were significant ( $\beta = 0.29$ ,  $CI_{95\%} = [0.0983, 0.4725]$ ,  $p < 0.05$ ). This result indicated that trust in chatbot and chatbot overload partially mediated the influence of chatbot competence on service evaluation.

### 4.3 Study C

Following the procedure in study B, study C was conducted from Aug 8th to Aug 23rd, 2023. Data were collected through a scenario-based questionnaire conducted on credamo.com, which includes five customer-chatbot dialogue scenarios (see Appendix III). A data sample of 263 participants was collected in Mainland China. After filling out the qualified responses using attention check

questions (e.g., *whether the chatbot used emoticons in dialogue?*), 232 ( $N_{\text{control}} = 46$ ,  $N_{\text{warmth}_1} = 46$ ,  $N_{\text{warmth}_2} = 48$ ,  $N_{\text{competence}} = 48$ ,  $N_{\text{hybrid}} = 44$ ) valid sample was used to further data analysis. To determine whether 232 responses were sufficient for the data analysis, we used the G\*Power 3.1.9.7 software to implement a post hoc statistical power analysis (Faul et al., 2007). The average effect size of the relationships presented in this research model was 0.22, with an alpha level of 0.05 and a power of 0.96, which was higher than the threshold value of 0.80. Thus, our sample size was sufficient for the data analysis. Table C1 presents the demographic profiles of the respondents. Among the 232 respondents, 66.80% were female, 29.70% were aged above 18 years and below 24 years, 83.60% had a bachelor's degree or below, 99.60% had more than two years of prior online shopping experience, 98.30% had more than two years of prior chatbot experience.

Similar to Study A and B, we used analyses of variance (ANOVAs) for the metric variables and Fisher's exact tests for the categorical variables. First, the levels of chatbot warmth ( $F(4, 227) = 15.96$ ,  $p = 0.00$ ) were found to differ significantly between the control and warmth\_1, warmth\_2, hybrid groups, the participants in the control group were likely to perceive that chatbot was warmer ( $M_{\text{control}} = 4.66$ ,  $SD = 1.29$ ) than those in the experimental groups ( $M_{\text{warmth}_1} = 6.09$ ,  $SD = 0.50$ ,  $p = 0.00$ ;  $M_{\text{warmth}_2} = 5.76$ ,  $SD = 0.80$ ,  $p = 0.00$ ;  $M_{\text{hybrid}} = 5.96$ ,  $SD = 0.78$ ,  $p = 0.00$ ). The levels of chatbot competence ( $F(4, 227) = 9.08$ ,  $p = 0.00$ ) were found to differ significantly between the control and competence, hybrid groups, the participants in the control group were likely to perceive that chatbot was more competent ( $M_{\text{control}} = 3.69$ ,  $SD = 1.29$ ) than those in the

<sup>5</sup> Warmth\_1 refers to chatbot with verbal warm cues, warmth\_2 refers to chatbot with verbal and non-verbal warm cues.

experimental groups ( $M_{\text{competence}} = 4.63$ ,  $SD = 1.62$ ,  $p = 0.00$ ;  $M_{\text{hybrid}} = 5.23$ ,  $SD = 1.17$ ,  $p = 0.00$ ). Therefore, our manipulation of chatbot warmth and chatbot competence was valid.

Second, the homogeneity of the conversation content was ensured because the participants reported no significant difference across the five experimental conditions regarding perceived informativeness ( $F(4, 227) = 2.22$ ,  $p = 0.07$ ). Moreover, the participants reported that they felt the experimental conditions were realistic and easy to understand, there was no significant difference across the five experimental conditions regarding perceived realism ( $F(4, 227) = 3.035$ ,  $p = 0.02$ ) and the participants reported only a minimal difference in perceived difficulty between the control and warmth groups ( $F(4, 227) = 3.53$ ,  $p = 0.01$ ). Third, the results showed that there was no significant group difference in gender ( $p = 0.62$ ), age ( $p = 0.09$ ), education ( $p = 0.70$ ), prior online shopping experience ( $p = 0.17$ ), prior chatbot experience ( $p = 0.49$ ). Thus, we concluded that the treatment was successful and the control variables would not confound the experimental effects.

As shown in Table C2, Cronbach's alpha ranged from 0.77–0.89, CR ranged from 0.78–0.94, and AVE ranged from 0.68–0.81, indicating satisfactory internal reliability. The item loadings ranged from 0.76–0.91, which indicated acceptable convergent validity. Moreover, discriminant validity was established because the square roots of the AVEs for all constructs were greater than the construct correlations shown in Table C3. Combined, these results indicate that the reliability and validity of the measurement model were satisfactory. The results of nonresponse bias and common method bias (see Appendix VI) showed that these biases were not serious in study C.

To determine the effects of chatbot warmth and chatbot competence, we compare cognitive trust, emotional trust, cognitive overload, and emotional overload in five experimental groups using ANOVA.

Our results indicated that there was significant difference across the experimental groups regarding cognitive trust ( $F(4, 227) = 2.73$ ,  $p = 0.03$ ) and emotional trust ( $F(4, 227) = 10.24$ ,  $p = 0.00$ ). Comparing to the control group, chatbot warmth and chatbot competence indicated a significant interaction effect on cognitive trust in hybrid group ( $M_{\text{control}} = 4.63$ ,  $SD = 1.35$ ,  $M_{\text{hybrid}} = 5.39$ ,  $SD = 1.19$ ,  $p = 0.00$ ). In contrast to control group, chatbot warmth indicated a significant main effect on emotional trust in warmth\_1 group ( $M_{\text{control}} = 4.52$ ,  $SD = 1.36$ ,  $M_{\text{warmth}_1} = 5.67$ ,  $SD = 0.73$ ,  $p = 0.00$ ), in warmth\_2 group ( $M_{\text{control}} = 4.52$ ,  $SD = 1.36$ ,  $M_{\text{warmth}_2} = 5.53$ ,  $SD = 1.14$ ,  $p = 0.00$ ), and in hybrid group ( $M_{\text{control}} = 4.52$ ,  $SD = 1.36$ ,  $M_{\text{hybrid}} = 5.53$ ,  $SD = 1.14$ ,  $p = 0.00$ ).

Moreover, our results indicated that there was significant difference across the experimental groups regarding cognitive overload ( $F(4, 227) = 3.89$ ,  $p = 0.00$ ) and emotional overload ( $F(4, 227) = 2.73$ ,  $p = 0.03$ ). Comparing to

the control group, chatbot warmth and chatbot competence indicated significant interaction effects on cognitive overload in hybrid group ( $M_{\text{control}} = 3.15$ ,  $SD = 0.99$ ,  $M_{\text{hybrid}} = 2.38$ ,  $SD = 1.10$ ,  $p = 0.05$ ), and emotional overload in hybrid group ( $M_{\text{control}} = 3.19$ ,  $SD = 1.42$ ,  $M_{\text{hybrid}} = 2.37$ ,  $SD = 1.15$ ,  $p = 0.03$ ).

## 5 Discussion and Implications

### 5.1 Key Findings

Although chatbots have been expected to replace humans and provide high-quality services for customers on e-commerce websites, unfavorable service evaluations and low conversion rates of chatbots have become challenges for online retailers. An anthropomorphic design has been proposed as a remedy to improve customers' beliefs of and behavioral intentions toward chatbot services. However, there is a limited amount of research on how chatbot anthropomorphism can best efficiently increase customers' purchase behaviors. Thus, this study aimed to examine the effects of chatbot anthropomorphic cues (i.e., warmth and competence) on customers' service evaluation and purchase behavior. We developed four versions of chatbots that differed in their degrees of warmth and competence but provided identical answers. The research model was tested using a pilot study ( $N = 33$ ), a lab-in-the-field experiment ( $N = 212$ ), and two scenario-based experiments ( $N = 124$ ,  $N = 232$ ). The empirical results support most hypotheses and presented several interesting findings.

First, both study A and study B showed that chatbot warmth and competence were positively related to trust in chatbot, which then increased service evaluation. Study C further verified that the interaction of chatbot warmth and chatbot competence helps increase cognitive trust and emotional trust. Prior studies on chatbots have found that chatbot anthropomorphism can increase customer trust (Cheng et al., 2022a, b; Pizzi et al., 2023) and trust in chatbot would facilitate customer attitude and acceptance of chatbots (Behera et al., 2021; Kyung & Kwon, 2022) in various service contexts. This study further established that chatbot anthropomorphism promoted customer service evaluation by increasing trust in chatbot. By humanizing chatbots with warmth and competence cues, this study showed a positive relationship between chatbot anthropomorphism and service evaluation in online shopping. Moreover, the findings showed that chatbot anthropomorphism facilitated service evaluation through the development of customer trust in chatbot. These findings indicate that customers are likely to show favorable attitudes toward chatbots that can make customers believe in their reliability and kindness.

Second, although the negative influence of chatbot overload on service evaluation was found supported in study A and study B, the effect of chatbot warmth on chatbot overload was found supported only in study A, and the effect of chatbot competence on chatbot overload was found supported only in study B. The result of study C provided a possible explanation that the co-existence of chatbot warmth and competence is necessary to alleviate both cognitive overload and emotional overload. Another plausible reason for the differential effects of chatbot warmth and competence on chatbot overload is customer gender, which has been proved to affect customers' evaluation of chatbot attributes when making purchase decisions (Sharma et al., 2022). Our additional analysis showed that gender positively moderated the effect of chatbot competence on chatbot overload in study A ( $\beta=0.39, p<0.01$ ) and study B ( $\beta=0.32, p<0.05$ ), which suggested that male customers are more likely to highly evaluate the competence aspect of chatbot in e-commerce service. However, gender didn't moderate the effect of chat warmth on chatbot overload, which encourages future research to explore the boundary condition that can explain when chatbot warmth helps effectively reduce chatbot overload. Additional analysis of study A showed that chatbot competence has a significant impact on emotional overload ( $\beta=-0.17, p<0.05$ ) but not cognitive overload ( $\beta=-0.01, p>0.05$ ), this also opens an avenue for future research to unpack when chatbot competence reduces cognitive overload. As an emerging technology in the service context, chatbot overload has been realized as a barrier that prevents users from effectively using chatbots (Kim et al., 2022; Lou et al., 2022; Wang et al., 2021). This study further demonstrated that the existence of chatbot overload makes customers have a negative evaluation of chatbot service and further showed that chatbot anthropomorphism helps alleviate chatbot overload in e-commerce service.

Third, service evaluation was found to have a significant effect on purchase intention and purchase behavior in study A. Our further analysis showed that the indirect effect of service evaluation on purchase behavior through purchase intention ( $\beta=0.84, CI_{95\%}=[0.5289, 1.2753]$ ) was significant, but the direct effects were nonsignificant ( $\beta=0.25, CI_{95\%}=[0.2543, -0.1762], p>0.05$ ). Thus, purchase intention fully mediated the effect of service evaluation on purchase behavior in online services provided by chatbots. That is, a customer's purchase intention could explain why increased service evaluation results in further purchase behavior. The findings are in line with the literature on the theory of planned behavior, which is a typical behavioral theory in information systems research, and argue that behavioral intention is the most proximal determinant of actual behavior (Leong et al., 2022). According to the theory of planned behavior (Ajzen, 1991), behavioral beliefs shape users' behavioral intentions, which in turn promotes

actual behaviors. Hence, the influence of service evaluation on purchase behavior is transmitted through customers' purchase intentions.

## 5.2 Theoretical Implication

This study contributes to the literature on chatbot anthropomorphism, chatbot user behavior, the Wixom & Todd model, and SIP in several ways. First, this study demonstrates that warmth and competence cues are effective in improving service evaluation and customer purchase, which advances our understanding of chatbot anthropomorphism in the e-commerce context. As an emerging self-service technology, the anthropomorphic chatbot has been increasingly implemented to replace manual services; however, the challenge of unfavorable evaluation and low conversion rate persists. Prior studies have provided fragmented evidence of the effects of chatbot anthropomorphism on service evaluation (e.g. Han et al., 2022) and customer purchase (e.g. Sharma et al., 2022), separately. Moreover, previous research has called for examining chatbot effectiveness in other service contexts characterised by different product categories and geographies (Behera et al., 2021; Sharma et al., 2022). This study extends existing work on chatbot anthropomorphism by examining a consolidated framework for explaining how chatbot warmth and competence contribute to an increase in service evaluation and customer purchase in the context of B2C e-commerce service.

Second, this study examines trust in chatbot and chatbot overload as parallel mechanisms in customer–chatbot interaction that can explain how chatbot anthropomorphism increases service evaluation. Prior studies have confirmed that trust in chatbot (Behera et al., 2021) and chatbot overload (Wang et al., 2021) affects customers' reactions to chatbots. This study further posited that trust in chatbot and chatbot overload could co-exist in the process of anthropomorphic chatbots serving customers. Based on the Wixom & Todd model and SIP, this study theorised trust in chatbot and chatbot overload as the behavioral beliefs that can be influenced by their object-based social beliefs of anthropomorphic chatbots and shape customer attitudes towards chatbot service. Thus, this study proposed a consolidated framework to explain how chatbot anthropomorphism influences service evaluation by changing customers' trust in chatbot and chatbot overload.

Third, guided by the Wixom & Todd model and SIP, this study contributes to the contextualisation of theoretical constructs in customer–chatbot communication. The Wixom & Todd model was initially employed to contextualise chatbot quality (Nguyen et al., 2022) and customer adoption of chatbots (von Walter et al., 2022). SIP was initially proposed to contextualise customer–chatbot relationship building (Hendriks et al., 2020), which can be influenced by chatbot

anthropomorphism in online service (Schanke et al., 2021). With the integration of the Wixom & Todd model and SIP, this study theories that chatbot warmth and chatbot competence represent object-based social beliefs, trust in chatbot, and chatbot overload represent behavioral beliefs, which sequentially influence service evaluation and customer purchase. Our findings, therefore, advance the application of the Wixom & Todd model and SIP in explaining customers' series reactions to chatbots. Based on these findings, future research is recommended to contextualise object-based attitudes and behavior-based traits which are also highlighted in the Wixom & Todd model (Taylor & Todd, 1995; Wixom & Todd, 2005) and SIP (Salancik & Pfeffer, 1978), to understand customer behavioral responses to anthropomorphic chatbots.

### 5.3 Practical Implication

This study has several implications for deriving the business value of chatbots by effectively managing customer beliefs and attitudes toward chatbots in online services. First, online retailers are recommended to develop an anthropomorphised chatbot that is characterised by warmth and competence traits. Because these traits are effective to cultivate customer trust in chatbot and relieve customers from the technology overload of using the chatbot. As an emerging online service technology, chatbots should be designed and configured to manage customer-chatbot interaction in online shopping. For instance, retailers are suggested to develop a chatbot that is characterised by warm cues (e.g., emoticons and short messages) to reassure customers who find unfamiliar technologies too intensive or confusing. A good-natured chatbot image could be tailored to relieve customers from the pressure caused by intensive, confusing, or intrusive messages and encourage them to maintain chatbot conversations frequently. Retailers could utilise a competent chatbot that conveys the cues of integrity, certainty, and confidence to reduce customers' uncertainty in online shopping. Displaying competence cues, such as rich information and detailed instructions, help decrease customer skepticism and develop the trustworthiness of chatbots in undertaking online services.

Second, it is recommended that online retailers should be simultaneously aware of the positive and negative influences of anthropomorphic chatbots on customer reactions to online services. Our findings indicate that both trust in chatbot and chatbot overload exert significant influences on service evaluation. Although anthropomorphic cues, such as automated responses and rich characters, make customers believe chatbots as humanlike agents, excessive and frequent messages would increase the difficulty of information processing and result in unfavorable service evaluation. Our additional analysis of study A showed that emotional overload ( $\beta = -0.27$ ,  $p < 0.001$ ) rather than cognitive overload ( $\beta = 0.06$ ,  $p > 0.05$ )

decreased service evaluation. Based on this finding, we recommend that retailers continue to improve the emotional intelligence of chatbots, such as disclosing their identities and communicating kindness and empathy in messages.

### 5.4 Limitations and Future Research

The limitations of this study offer opportunities for future research. First, this study focused on examining the effects of users' object-based social beliefs and behavioral beliefs in predicting service evaluation and customer purchase. Prior studies have examined the influence of customer attitudes on predicting customers' reactions to chatbots (e.g., Behera et al., 2021). The Wixom & Todd model also emphasises the role of object-based attitudes, which were not considered in this study in its examination of how customers' object-based beliefs influence their behavioral attitudes and behavioral intentions. Therefore, future research could extend our research model by incorporating the roles of object-based attitude (e.g., information satisfaction) in facilitating the contextualisation of the Wixom & Todd model in online service chatbots.

Second, although the experimental design used in this study helped increase internal validity, the small sample size and homogenous demographic backgrounds (e.g., gender, age, and education) may limit the generalizability of the research findings. Therefore, future research should be based on a large sample of chatbot users with diverse demographic characteristics derived from data collected from market research institutions. Moreover, empirical datasets collected from different countries and regions would help evaluate the effects of socioeconomic factors on customers' reactions to chatbots.

Third, in this study, unexpected findings were that the effects of chatbot warmth and competence on chatbot overload differ in terms of customer demographics. Although the moderating effect of customer gender provides a plausible explanation for this result, future research should be conducted to explain this surprising finding based on additional theoretical or empirical evidence. A plausible direction is to extend the context of chatbot service, which may moderate the effects of chatbot warmth and competence on chatbot overload. In prior research, service context and chatbot type have been considered possible contingencies that regulate customers' responses to anthropomorphic chatbots (Blut et al., 2021). For instance, product type (e.g., search vs. experience) (Zhu et al., 2022) or chatbot interface (text vs. speech) (Rzepka et al., 2022) may serve as potential boundary conditions for the effects of chatbot warmth and competence on chatbot overload. Future research could further verify the research model by considering the moderating effects of search products and speech-based chatbots.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10796-023-10438-x>.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (No. 72002062, 72332007), Fundamental Research Funds for the Central Universities (No. JZ2023H-GTB0277), Hunan Provincial Natural Science Foundation of China (No. 2022JJ40655).

## Declarations

**Competing Interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- Ahmad, R., Siemon, D., Gnewuch, U., & Robra-Bissantz, S. (2022). Designing personality-adaptive conversational agents for mental health care. *Information Systems Frontiers*, 24(3), 923–943. <https://doi.org/10.1007/s10796-022-10254-9>
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- 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, 183–189. <https://doi.org/10.1016/j.chb.2018.03.051>
- Behera, R. K., Bala, P. K., & Ray, A. (2021). Cognitive chatbot for personalised contextual customer service: Behind the scene and beyond the hype. *Information Systems Frontiers*, 1–22. <https://doi.org/10.1007/s10796-021-10168-y>
- Blut, M., Wang, C., Wunderlich, N. V., & Brock, C. (2021). Understanding anthropomorphism in service provision: A meta-analysis of physical robots, chatbots, and other AI. *Journal of the Academy of Marketing Science*, 49(4), 632–658. <https://doi.org/10.1007/s11747-020-00762-y>
- Cenfetelli, & Bassellier. (2009). Interpretation of formative measurement in information systems research. *MIS Quarterly*, 33(4), 689. <https://doi.org/10.2307/20650323>
- Chan, E., & Ybarra, O. (2002). Interaction goals and social information processing: Underestimating one's partners but overestimating one's opponents. *Social Cognition*, 20(5), 409–439. <https://doi.org/10.1521/soco.20.5.409.21126>
- Chen, X., Wei, S., & Rice, R. E. (2020). Integrating the bright and dark sides of communication visibility for knowledge management and creativity: The moderating role of regulatory focus. *Computers in Human Behavior*, 111, 106421. <https://doi.org/10.1016/j.chb.2020.106421>
- Cheng, X., Bao, Y., Zarifis, A., Gong, W., & 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), 496–517. <https://doi.org/10.1108/INTR-08-2020-0460>
- Cheng, X., Zhang, X., Cohen, J., & Mou, J. (2022b). Human vs. AI: Understanding the impact of anthropomorphism on consumer response to chatbots from the perspective of trust and relationship norms. *Information Processing & Management*, 59(3), 102940. <https://doi.org/10.1016/j.ipm.2022.102940>
- Chien, S.-Y., Lin, Y.-L., & Chang, B.-F. (2022). The effects of intimacy and proactivity on trust in human-humanoid robot interaction. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-022-10324-y>
- Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement*, 12(4), 425–434. <https://doi.org/10.1177/014662168801200410>
- 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), 132–148. <https://doi.org/10.1177/00222429211045687>
- Davis, J. M., & Agrawal, D. (2018). Understanding the role of interpersonal identification in online review evaluation: An information processing perspective. *International Journal of Information Management*, 38(1), 140–149. <https://doi.org/10.1016/j.ijinfomgt.2017.08.001>
- Ehrke, F., Bruckmüller, S., & Steffens, M. C. (2020). A double-edged sword: How social diversity affects trust in representatives via perceived competence and warmth. *European Journal of Social Psychology*, 50, 1540–1554. <https://doi.org/10.1002/ejsp.2709>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6), 878–902. <https://doi.org/10.1037/0022-3514.82.6.878>
- Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research*, 18(3), 382. <https://doi.org/10.2307/3150980>
- Fu, S., Li, H., Liu, Y., Pirkkalainen, H., & Salo, M. (2020). Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload. *Information Processing & Management*, 57(6), 102307. <https://doi.org/10.1016/j.ipm.2020.102307>
- George, D., & Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge. <https://doi.org/10.4324/9780429056765>
- Go, E., & Sundar, S. S. (2019). Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. *Computers in Human Behavior*, 97, 304–316. <https://doi.org/10.1016/j.chb.2019.01.020>
- Gong, T., Choi, J. N., & Samantha, M. (2016). Does customer value creation behavior drive customer well-being? *Social Behavior and Personality*, 44(1), 59–75. <https://doi.org/10.2224/sbp.2016.44.1.59>
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (6th ed.). Prentice-Hall International Inc. <http://www.sciepub.com/reference/219114>. Accessed 10 Oct 2022.
- Hair, J. F., Jr., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.
- Han, E., Yin, D., & Zhang, H. (2022). Bots with feelings: Should AI agents express positive emotion in customer service? *Information Systems Research*, 1–16. <https://doi.org/10.1287/isre.2022.1179>
- Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, 24(10), 1918–1927. <https://doi.org/10.1177/0956797613480187>
- Hendriks, F., Ou, C. X. J., Khodabandeh Amiri, A., & Bockting, S. (2020). The power of computer-mediated communication theories in explaining the effect of chatbot introduction on user experience.



- Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2020.034>
- Hildebrand, C., & Bergner, A. (2021). Conversational robo advisors as surrogates of trust: Onboarding experience, firm perception, and consumer financial decision making. *Journal of the Academy of Marketing Science*, 49(4), 659–676. <https://doi.org/10.1007/s11747-020-00753-z>
- Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science*, 10(6), 791–815.
- Jiang, Y. (2023). Make chatbots more adaptive: Dual pathways linking human-like cues and tailored response to trust in interactions with chatbots. *Computers in Human Behavior*, 138, 107485. <https://doi.org/10.1016/j.chb.2022.107485>
- Keh, H. T., & Sun, J. (2018). The differential effects of online peer review and expert review on service evaluations: The roles of confidence and information convergence. *Journal of Service Research*, 21(4), 474–489. <https://doi.org/10.1177/1094670518779456>
- Kim, G., Shin, B., & Lee, H. G. (2009). Understanding dynamics between initial trust and usage intentions of mobile banking. *Information Systems Journal*, 19(3), 283–311. <https://doi.org/10.1111/j.1365-2575.2007.00269.x>
- Kim, J. H., Kim, M., Kwak, D. W., & Lee, S. (2022). Home-tutoring services assisted with technology: Investigating the role of artificial intelligence using a randomized field experiment. *Journal of Marketing Research*, 59(1), 79–96. <https://doi.org/10.1177/00222437211050351>
- Kumar, V., Rajan, B., Salunkhe, U., & Joag, S. G. (2022). Relating the dark side of new-age technologies and customer technostress. *Psychology & Marketing*, 39(12), 2240–2259. <https://doi.org/10.1002/mar.21738>
- Kyung, N., & Kwon, H. E. (2022). Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. *Production and Operations Management*, 1–20. <https://doi.org/10.1111/poms.13785>
- Leong, L.-Y., Hew, T.-S., Ooi, K.-B., Metri, B., & Dwivedi, Y. K. (2022). Extending the theory of planned behavior in the social commerce context: A Meta-Analytic SEM (MASEM) Approach. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-022-10337-7>
- Lewis, B. R., Templeton, G., & Byrd, T. (2005). A methodology for construct development in MIS research. *European Journal of Information Systems*, 14(4), 388–400. <https://doi.org/10.1057/palgrave.ejis.3000552>
- Li, L., Lee, K. Y., Emokpae, E., & Yang, S.-B. (2021). What makes you continuously use chatbot services? Evidence from chinese online travel agencies. *Electronic Markets*, 31(3), 575–599. <https://doi.org/10.1007/s12525-020-00454-z>
- Lou, C., Kang, H., & Tse, C. H. (2022). Bots vs. humans: how schema congruity, contingency-based interactivity, and sympathy influence consumer perceptions and patronage intentions. *International Journal of Advertising*, 41(4), 655–684. <https://doi.org/10.1080/02650487.2021.1951510>
- Lowry, P. B., & Gaskin, J. (2014). Partial least squares (pls) structural equation modeling (sem) for building and testing behavioral causal theory: When to choose it and how to use it. *IEEE Transactions on Professional Communication*, 57(2), 123–146. <https://doi.org/10.1109/TPC.2014.2312452>
- Lu, J., Zhang, Z., & Jia, M. (2019). Does servant leadership affect employees' emotional labor? A social information-processing perspective. *Journal of Business Ethics*, 159(2), 507–518. <https://doi.org/10.1007/s10551-018-3816-3>
- Luo, X., Qin, M. S., Fang, Z., & Qu, Z. (2021). Artificial intelligence coaches for sales agents: Caveats and solutions. *Journal of Marketing*, 85(2), 14–32. <https://doi.org/10.1177/0022242920956676>
- Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, 13. <https://doi.org/10.1177/002224379102800302>
- Mostafa, R. B., & Kasamani, T. (2022). Antecedents and consequences of chatbot initial trust. *European Journal of Marketing*, 56(6), 1748–1771. <https://doi.org/10.1108/EJM-02-2020-0084>
- Moussawi, S., Koufaris, M., & Benbunan-Fich, R. (2021). How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents. *Electronic Markets*, 31(2), 343–364. <https://doi.org/10.1007/s12525-020-00411-w>
- Nguyen, Q. N., Ta, A., & Prybutok, V. (2019). An integrated model of voice-user interface continuance intention: The gender effect. *International Journal of Human-Computer Interaction*, 35(15), 1362–1377. <https://doi.org/10.1080/10447318.2018.1525023>
- Nguyen, T., Quach, S., & Thaichon, P. (2022). The effect of AI quality on customer experience and brand relationship. *Journal of Consumer Behaviour*, 21(3), 481–493. <https://doi.org/10.1002/cb.1974>
- Nunnally, J. C. (1978). *Psychometric methods*. McGraw-Hill.
- Petter, Straub, & Rai. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, 31(4), 623. <https://doi.org/10.2307/25148814>
- Pizzi, G., Vannucci, V., Mazzoli, V., & Donvito, R. (2023). I, chatbot! the impact of anthropomorphism and gaze direction on willingness to disclose personal information and behavioral intentions. *PSychology & Marketing*, 40(7), 1372–1387. <https://doi.org/10.1002/mar.21813>
- Reinartz, W. J., Haenlein, M., & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM. *Social Science Research Network*, 26(4), 332–334. <https://doi.org/10.1016/j.ijresmar.2009.08.001>
- Rhim, J., Kwak, M., Gong, Y., & Gweon, G. (2022). Application of humanization to survey chatbots: Change in chatbot perception, interaction experience, and survey data quality. *Computers in Human Behavior*, 126, 107034. <https://doi.org/10.1016/j.chb.2021.107034>
- Roccapriore, A. Y., & Pollock, T. G. (2023). I don't need a degree, I've got abs: Influencer warmth and competence, communication mode, and stakeholder engagement on social media. *Academy of Management Journal*, 66(3), 979–1006. <https://doi.org/10.5465/amj.2020.1546>
- Roy, R., & Naidoo, V. (2021). Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research*, 126, 23–34. <https://doi.org/10.1016/j.jbusres.2020.12.051>
- Rutkowski, A., Saunders, C., & Wiener, M. (2013). Intended usage of a healthcare communication technology: Focusing on the role of it-related overload. *International Conference on Information Systems*, 17. <https://www.researchgate.net/publication/348372898>. Accessed 10 Oct 2022.
- Rzepka, C., Berger, B., & Hess, T. (2022). Voice Assistant vs. Chatbot – Examining the fit between conversational agents' interaction modalities and information search tasks. *Information Systems Frontiers*, 24(3), 839–856. <https://doi.org/10.1007/s10796-021-10226-5>
- Salancik, G. R., & Pfeffer, J. (1978). A social information processing approach to job attitudes and task design. *Administrative Science Quarterly*, 23(2), 224. <https://doi.org/10.2307/2392563>
- Saunders, C., Wiener, M., Klett, S., & Sprenger, S. (2017). The impact of mental representations on ICT-related overload in the use of mobile phones. *Journal of Management Information Systems*, 34(3), 803–825. <https://doi.org/10.1080/07421222.2017.1373010>
- Schanke, S., Burtch, G., & Ray, G. (2021). Estimating the impact of “humanizing” customer service chatbots. *Information Systems Research*, 32(3), 736–751. <https://doi.org/10.1287/isre.2021.1015>

- Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, 37(3), 875–900. <https://doi.org/10.1080/07421222.2020.1790204>
- Seeger, A.-M., Pfeiffer, J., & Heinzl, A. (2021). Texting with human-like conversational agents: Designing for anthropomorphism. *Journal of the Association for Information Systems*, 22(4), 931–967. <https://doi.org/10.17705/1jais.00685>
- Sharma, M., Joshi, S., Luthra, S., & Kumar, A. (2022). Impact of digital assistant attributes on millennials' purchasing intentions: A multi-group analysis using PLS-SEM, Artificial Neural Network and fsQCA. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-022-10339-5>
- Shen, X.-L., Li, Y.-J., Sun, Y., & Wang, N. (2018). Channel integration quality, perceived fluency and omnichannel service usage: The moderating roles of internal and external usage experience. *Decision Support Systems*, 109, 61–73. <https://doi.org/10.1016/j.dss.2018.01.006>
- Sun, Y., Li, S., & Yu, L. (2022). The dark sides of AI personal assistant: Effects of service failure on user continuance intention. *Electronic Markets*, 32(1), 17–39. <https://doi.org/10.1007/s12525-021-00483-2>
- Taylor, S., & Todd, P. (1995). Assessing IT usage: The role of prior experience. *MIS Quarterly*, 19(4), 561. <https://doi.org/10.2307/249633>
- von Walter, B., Kremmel, D., & Jäger, B. (2022). The impact of lay beliefs about AI on adoption of algorithmic advice. *Marketing Letters*, 33(1), 143–155. <https://doi.org/10.1007/s11002-021-09589-1>
- Wang, W., Chen, L., Xiong, M., & Wang, Y. (2021). Accelerating AI adoption with responsible AI signals and employee engagement mechanisms in health care. *Information Systems Frontiers*. <https://doi.org/10.1007/s10796-021-10154-4>
- Weisband, S. P., Schneider, S. K., & Connolly, T. (1995). Computer-mediated communication and social information: Status salience and status differences. *Academy of Management*, 38(4), 1124–1151. <https://doi.org/10.2307/256623>
- Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, 16(1), 85–102.
- Xu, J. (David), Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: An empirical test in the e-service context. *MIS Quarterly*, 37(3), 777–794. <https://doi.org/10.25300/MISQ/2013/37.3.05>
- Zhu, Y., Zhang, J., Wu, J., & Liu, Y. (2022). AI is better when I'm sure: The influence of certainty of needs on consumers' acceptance of AI chatbots. *Journal of Business Research*, 150, 642–652. <https://doi.org/10.1016/j.jbusres.2022.06.044>
- Zogaj, A., Mähner, P. M., Yang, L., & Tschulin, D. K. (2023). It's a Match! The effects of chatbot anthropomorphization and chatbot gender on consumer behavior. *Journal of Business Research*, 155, 113412. <https://doi.org/10.1016/j.jbusres.2022.113412>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

**Yang Li** is a lecturer at School of Management, Hefei University of Technology. She received her PhD candidate in Information Systems at the joint PhD program between the University of Science and Technology of China and City University of Hong Kong. Her research interests include omnichannel management, supply chain management, and e-commerce. She has published in academic journals including *Decision Support Systems*, *IEEE Transactions on Engineering Management*, *Journal of Business Research*, and etc.

**Zhenghua Gan** is a postgraduate at School of Management, Hefei University of Technology. Her research interests include artificial intelligence for business.

**Bowen Zheng** is currently an Assistant Professor at Business School, Central South University. She earned her PhD degree with the University of Science and Technology of China and City University of Hong Kong. Her research interests include computer-mediated-communication, digital workplace and consumer behavior. She has published in *Information & Management*, *Internet Research*, *Information Technology & People*, *Public Relations Review* and in the academic conference PACIS.