Title page

Classifying and measuring the service quality of AI Chatbot in frontline service

Qian Chen[a], Yeming Gong[b], Yaobin Lu[c], Jing Tang[d]

Qian Chen

-Email address: chen_qian@mail.hzau.edu.cn
-Mailing address: College of economics & management, Huazhong Agricultural University, Wuhan, China.

Yeming Gong

-Email address: gong@em-lyon.com
-Mailing address: EMLYON Business School, Ecully Cedex 69134, France
-Tel:+33-4-78337742, Fax:+33-4-78337928

Yaobin Lu (Corresponding author)

-Email:luyb@ hust.edu.cn
-Mailing address: School of Management, Huazhong University of Science and Technology, Wuhan, China.
-Phone:+86-27-87556448

Jing Tang

-Email address: jtang@saunders.rit.edu
-Mailing address: Saunders College of Business, Rochester Institute of Technology, 105 Lomb Memorial Drive, LOW-A314, Rochester, NY 14623, USA
-Tel:+1-585-475-2015

Classifying and measuring the service quality of AI Chatbot in frontline service

ABSTRACT

AI chatbots have been widely applied in the frontline to serve customers. Yet, the existing dimensions and scales of service quality can hardly fit the new AI environment. To address this gap, we define the dimensions of AI chatbot service quality (AICSQ) and develop the associated scales with a mixed-method approach. In the qualitative analysis, with the coding of the interviews from 55 global organizations in 17 countries and 47 customers, we develop new multi-level dimensions of AICSQ, including seven second-order and 18 first-order constructs. Then we follow a 10-step scale development method to establish the valid scales. The nomological test result shows that AICSQ positively influences customers' satisfaction with, perceived value of, and intention of continuous use of AI chatbots. The innovative dimensions and scales of AI chatbot service quality provide conceptual classification and measurement instruments for the future study of chatbot service in the frontline.

Keywords: AI chatbot; service quality; artificial intelligence; scale development; mixed-method approach.

1. Introduction

A growing number of organizations adopt artificial intelligence (AI) chatbots to serve customers in the frontline (De Keyser et al., 2019). Gartner (2019) shows that by 2021, 15% of all customer service interactions globally will be completely handled by AI, which is an increase of 400% from 2017

(Gartner, 2019). Besides, according to the "2019 China Artificial Intelligence Industry Research Report" released by the China Economic and Information Research Center in December 2019, the scale of AI chatbot business reached 2.72 billion yuan in 2018, and it is expected that the number will exceed 16 billion yuan in 2022 (China Economic and Information Research Center, 2019). The new frontline service providers of AI chatbots are different from human customer service as well as self-service technology (SST). Combining advanced technologies that include natural language processing (NLP), cloud, machine learning, and biometrics, AI chatbots not only provide customers immediate and consistent services but also reduce the operating costs for organizations (Deloitte, 2018a; Wirtz et al., 2018). Previous studies already show that, in traditional service, the service quality of frontline employees who directly contact customers will significantly affect customer satisfaction, customer loyalty, and organization profits (Arora & Narula, 2018; Cenfetelli et al., 2008; Maklan et al., 2017; Piercy, 2014). Despite the growing attention to AI service quality in practice and academics (Jörling et al., 2019; Luo et al., 2019; Thomaz et al., 2020), the conception of AI chatbot service quality and the related evaluation instruments are still unclear (De Keyser et al., 2019), leading to the lack of theoretical perspective to evaluate and improve AI chatbot service quality (AICSQ). Since the conception, dimension, and measurement are vital to the research on service quality, many researchers study the classification and scale of service quality. A widely-used scale of service quality is SERVQUAL proposed by Parasuraman et al. (1988). This scale is verified in various online and offline contexts. Other researchers also developed service quality dimensions and scales specifically for online shopping, E-service, or information systems, such as Zeithaml et al. (2000), Cox & Dale (2001), Wolfinbarger & Gilly(2003), Collier & Bienstock (2006), Xu et al.(2013), and Tan et al.(2017).

However, these dimensions and scales of service quality are challenged when applied to the context of AI chatbot service. First, compared with human customer service, the AI chatbot service is based on an advanced information system. The existing dimensions of human customer service quality, such as tangibles, reliability, responsiveness, assurance, and empathy, cannot fully reflect the innovative technical characteristics of AI chatbot service. Second, compared with websites or information systems, the AI chatbot has human-like intelligence, such as language intelligence and recognition capabilities, which are not shown in the existing studies. This study theoretically classifies and develops scales of AICSQ, which provide a framework and benchmark for evaluating, designing, and improving AI chatbot service quality for practitioners. To extend the service quality research to consider the quality evaluation in the AI environment, the research questions of this study are: (1) What are the dimensions of AI chatbot service quality? (2) How to measure AI chatbot service quality?

To build a validated measurement model of AI chatbot service quality, we adopted a mixed-method approach, integrating qualitative and quantitative methods. The qualitative section aims to classify the dimensions of AICSQ. We used the coding technique of grounded theory to analyze the semi-structured interview data, which are collected from both organizations and customers. Then we adopted a quantitative method to develop the AICSQ scale and tested the discriminant, convergent, predictive validity, and reliability of the measurement model. We obtained an AICSQ scale with good reliability and validity through the three-round of data collections and statistical tests: pretest, refine and validation, and nomological test.

Our research contributes to service quality studies, AI service science, and practical applications in business. Theoretically, the multi-dimensional measures of AICSQ solve the problem that the

existing service quality dimension cannot comprehensively reflect the AI chatbot service context. Empirically, we introduce new validated scales for AICSQ, providing a measurement instrument for future empirical studies on AI chatbot service, especially the role of AI chatbot service quality in customer satisfaction and business outcomes. In practice, the dimension and scale of AICSQ we developed also play an important role in AI chatbot development phases, including investigating, designing, developing, evaluating, and customer relationship maintenance.

2. Theoretical background

2.1 Concept, dimensions, and measurement of service quality

From service creation to service delivery, the service process involves two types of subjects: service providers (organizations) and service users (customers). Regarding the definition of service quality, from the perspective of a customer, service quality is a perception and is an overall subjective evaluation of the service in the interactive process of service delivery (Dabholkar et al., 2000; Parasuraman et al., 1988). From an organization's perspective, service quality is evaluated by the extent to which it is consistent with the initial design (Golder et al., 2012). For organizations, providing the service that conforms to the blueprint is the basis of ensuring customer satisfaction. In this case, the service quality of AI chatbots depends on organizations' AI technology applications and service design. For AI chatbots service, organizations also attach importance to customers' feedback so that they can continuously iterate and improve their service. In an AI environment, the periods of iteration are significantly reduced, and organizations create service in a more integrative, heterogeneous, and flexible way (Jörling, Böhm, & Paluch, 2019). Thus, integrating both organization and customer perspectives is necessary and important in identifying service quality in an AI service environment.

Yet, we find most existing studies on service quality are conducted under a single perspective and the studies on service quality under AI context are under researched. To fill this gap, this study examines the service quality of AI chatbot from the perspectives of both organizations and customers. Specifically, we develop the dimensions of AI chatbot service quality based on the interview data from AI chatbot designers (organizations) and customers.

Service quality is a multi-dimensional concept (Korfiatis et al., 2019; Mittal et al., 1999). Scholars have explored the dimensional composition and measurement scales of service quality in different contexts. One of the widely used classifications and scales is SERVQUAL proposed by Parasuraman et al.(1988). This scale has been widely used not only in offline contexts but also in online environments (Jiang et al., 2002; O'Neill et al., 2001). However, there are significant differences between online and offline services. For example, in the online context, the users interact with the website as well as the service providers (Cai & Jun, 2003). While in the offline context, the users only interact with the service providers. An increasing number of studies have explored the dimensions and scales of service quality with the aim to measure it more concisely and accurately. Among them, most studies focus on service quality in online shopping settings, such as the CiteQUAL (Yoo & Donthu, 2001), the WEBQUAL 1.0 to WEBQUAL 4.0 (Barnes & Vidgen, 2001; Barnes & Vidgen, 2002), the eTailQ (Wolfinbarger & Gilly, 2003), the 11 dimensions of the service quality of online shopping proposed by Zeithaml et al. (2005), the four dimensions proposed by Cox & Dale (2001), and the dimension summarization in review work (Madu & Madu, 2002). These studies capture the features of online service quality. However, these studies are conducted from only one perspective of one player on the platform. Some scholars have already realized the shortages of measuring from only one

perspective and started to develop the scales of service quality from different players' perspectives. For example, Cai & Jun (2003) suggested the four-element dimension of e-service quality, including website design/content, trustworthiness, prompt/reliable service, and communication, with an integrated perspective of both online buyers and information searchers. Parasuraman, Zeithaml, & Malhotra (2005) reviewed the research on e-service service quality in those years. After analyzing the shortcomings of the existing scales such as WebQual, SITEQUAL, eTailQ in measuring E-service, they propose E-RecS-QUAL, including 11 items in three dimensions. In addition to the above mentioned classic studies on dimensions of E-service quality, many other studies contribute to the classification and measurement of service quality in specific e-context, such as government website service quality (Tan et al., 2013). However, these service quality dimensions are not applicable enough in the frontline AI service because it provides more integrative, heterogeneous, and flexible service. Supported with AI technology, frontline AI chatbot service shows new features. For example, humanoid interaction runs through the whole service process, and service quality can be improved in short periods due to its self-learning ability. To address this gap, this work applies an integrated perspective to identify the dimensions of AICSQ, which can reflect the new features of service in a frontline AI environment.

2.2 AI-service quality

AI chatbot service is not only a new kind of service provider but also an innovative service method. AI chatbot is different from traditional online virtual avatars, which have real human customer service behind them. Also, AI chatbot service is different from SST based on information systems. As a new species, AI chatbot is anthropomorphic, but it surpasses humans in certain aspects of intelligence, such

as information storage, computing power, and learning ability. In some aspects, it is temporarily inferior to human beings, such as emotional intelligence (Gray et al., 2007). AI chatbot has its unique inherent characteristics, service methods, and service output content. The difference in service quality between AI chatbot, human customer service, and SST is shown in Table 1, partially based on Wirtz et al. (2018).

————————————————

Insert Table 1 here

————————————————

Table 1 suggests that compared with SST, the AI chatbot is more flexible. Compared with human customer service, some capabilities of AI chatbots surpass that of humans, such as self-learning, accurate personalized recommendation, being always online, ready to respond to customers' requirements anytime, and is more stable (Deloitte, 2018b). While some capabilities of AI chatbots are not as good as humans, such as the inability to have deep emotional interaction with customers and some difficult-to-solve problems still require human customer service (Forbes, 2017). Organizations benefit from AI chatbots and have the motivation to promote AI chatbot services. Now, AI chatbot has been widely used in e-commerce, and AI chatbot is often the first "employee" for consumers to contact the organization. Many customers have to first interact with AI chatbots before conducting purchase behavior. Only when the customers encounter a problem that AI chatbot cannot solve do they transfer to the human customer service. Therefore, the quality of AI chatbots will impact customers' first impression of and satisfaction with the organization.

The existing studies mainly investigate customers' interaction with frontline AI chatbots in two streams. One stream is to study customers' experiences and behaviors in the interaction process. Another is to study the consequences of customer-AI chatbot interaction. From the perspective of the

interaction process, the researchers find that customers are interested in interacting with AI chatbot and do not find an uncanny valley effect in the human-AI chatbot interaction context (Ciechanowski et al., 2019). Researchers also compare users' behavior when they interact with humans and that when with AI chatbots. They find that when users interact with a human being, they are "more open, more conscientious, more extroverted, more agreeable, and self-disclosing than AI"(Mou & Xu, 2017, p432). Novak & Hoffman (2019) find that customers' experience of smart objects includes self-expansion, self-extension, self-restriction, and self-reduction when exploring the relationship between consumers and smart objects. From the perspective of the interacting consequences, researchers study the effects of customers' satisfaction with, acceptance of, continuous use of, and trust in AI chatbots (Loureiro et al., 2021). The information quality, service quality, perceived enjoyment, perceived usefulness, and perceived ease of use affect customers' satisfaction and continuous use (Ashfaq et al., 2020). Customers' attachment and perceived value of AI chatbots also positively affect customers' satisfaction, commitment, and trust in AI chatbots (Loureiro et al., 2021). Besides, Chung et al. (2018) find that using AI chatbots can improve customers' satisfaction with the brand. The above studies show AI chatbot's values to both customers and organizations. Improving frontline service quality will enhance customers' satisfaction and loyalty. Since AI chatbots have become important frontline workers, enhancing AI chatbot service quality is important for organizations.

3. AI chatbot service quality (AICSQ) scale development

The AICSQ scale development took three phases. In phase 1, we developed the multiple-level dimension of AICSQ by adopting the coding method of grounded theory, which is suitable for studying under-researched and emerging phenomena (Corbin & Strauss, 1990). Phase 2 and phase 3 are the

quantitative sections of AICSQ scale development. Following the ten-step method of scale development (Mackenzie et al., 2011), we obtained AICSQ with reliability and validity in phase 2, then conducted a nomological test in phase 3 by further constructing and examining a structure model. Figure 1 shows the roadmap of the research design.

—————————————————

Insert Figure 1 here

—————————————————

3.1 Phase 1: Developing the multi-level dimensions of AI chatbot service quality

The aim of phase 1 is to develop dimensions of AICSQ. The developing process was divided into four stages. Following the coding process of Thomas & Bostrom (2010) and the coding method of Corbin & Strauss (1990), in the first stage, we made a coding scheme based on the existing literature, which included the concepts related to service quality and the characteristics of AI chatbot service. On this basis, we randomly chose 10% of the samples for double-blind independent pre-coding. The purpose of pre-coding is to ensure the reliability of the coding results. In the process, two coders first compared and discussed the pre-coding results and then modified the coding scheme according to the pre-coding results. In the second stage of open coding, two coders independently read the interview text sentence by sentence, extracting concepts related to service quality. Then we calculated the inter-coder reliability, computed at 91%, indicating that the conformity between the coders is acceptable (Weber, 1990). In stage three, the two coders independently conducted axial coding based on the results of open codes. Then we invited two outside experts to discuss the results until we reached a consensus. In stage four, we conducted selective coding to cluster the axial codes obtained in stage three. The two coders independently performed selective coding, then discussed with experts repeatedly until they reached

9

100% consensus, and finally formed seven selective codes, which were second-order constructs in the multi-level dimension.

3.1.1 Data collection

To study the AI chatbot service quality from a comprehensive perspective, we collected a total of 1176 pages of interview data from both organizations and customers. (1) Data collection from organization sources: We conducted semi-structured questionnaire interviews with 55 global organizations that use AI chatbot service in 17 countries (France, USA, UK, Sweden, India, China, Japan, Switzerland, Denmark, Norway, The Netherlands, Germany, Italia, Spain, Australia, Austria, and Belgium). The organizations are in different industries, including hotel service, travel, food service, finance service, healthcare, consulting, retailing, entertainment, and career & education. The data collection lasted from December 2018 to May 2020. We finally obtained 1136 pages of textual data. The position of interviewees we selected are mainly four types: CEO or Vice presidents, senior IT Engineers, IT service managers, and marketing managers, since they were the stakeholders of AI chatbot service in organizations. The outline of the semi-structured questionnaire is shown in Appendix A. We hired ten research assistants to conduct the interview in Europe. The language used in the questionnaire and interview was English. (2) Data collection from customer sources: After a preliminary survey, we found that university students did online shopping and interacted with AI chatbots frequently. Therefore, we recruited students from a university in China to conduct focus group interviews. We held five focus group discussions and 47 students in total participated in the discussions. We asked them about their experience of and opinions on AI chatbots service. Finally, we recorded 40 pages of notes.

3.1.2 The results of text-coding

Through open coding, axial coding, and selective coding, we established a seven-dimension classification of AICSQ. There are two or three sub-dimensions subordinate to each dimension, respectively. The multi-level dimension of AICSQ is shown in Figure 2.

—————————————————

Insert Figure 2 here

—————————————————

Semantic understanding: Customers' demand for customer service is to solve their requirements. For AI chatbot service, the prerequisite for solving requirements is that the AI chatbot can understand the requirements sent by customers in a voice or textual way. Customers' judgments on the level of an AI chatbot's understanding ability are based on the accuracy of their feedback. From the text data, we find that customers not only require AI chatbot to understand the content of the queries but also hope that AI chatbot can understand their emotional expressions in the conversation. Based on understanding customers' requirements and emotion, AI chatbot can give appropriate responses to them. The text examples are shown in Table 2.

Close Human-AI Collaboration: AI and human customer service have their respective advantages. Compared with humans, AI chatbots are superior in calculation ability. Human customer service is good at solving specific personalized problems, which especially are involved in deep emotional interaction (Deloitte, 2018a). The two types of customer service can achieve complementary advantages. The interview data also show that AI chatbots need to cooperate with manual workers to achieve a high-level service quality when handling complex requirements. Some interviewees mentioned that AI chatbots are not yet smart enough. Some services may be too complicated for the machine to completely meet customers' needs without human operation (Ba et al., 2010). Sometimes,

11

the traditional barrier exists in an AI chatbot adoption, that is, customers have a habit or need to interact with humans (Mani & Chouk, 2018). The textual data in Table 2 show that, for customers, their perceived close human-AI collaboration is reflected in three aspects: first, customers need easy access to obtain desired services, e.g. they can get both AI and human customer service on the interface easily. Second, customers want a quick switch between AI and human service when necessary. Specifically, when AI chatbots cannot solve the problem, customers need to be able to switch to human customer service quickly; conversely, when customers need the quick service of AI chatbots, they also need to be able to switch from human worker to AI chatbot service easily. Third, when a requirement that cannot be solved is handed over to human customer service, customers do not need to repeat their requirement, and human-AI collaboration is required to be memorable.

Human-like: The long-term human customer service makes customers have adapted to them. Some customers claim that they prefer interacting with human beings (Mani & Chouk, 2018) and feel more comfortable when a human serves them (Ba et al., 2010). When interacting with AI chatbots, customers expect they have human-like characteristics, which are also an important aspect that reflects the intelligence of AI services (Tian et al., 2017). Previous studies show that a reason for customers' rejection of SST is that they prefer interpersonal interaction (Collier & Kimes, 2012; Lee & Lyu, 2016). AI chatbot adoption in the frontline alleviates the problem (Sheehan et al., 2020). Customers' human-like experience of AI chatbot enhances firm-customer interactions (Murtarelli et al., 2021) and makes their evaluation of AI chatbots more positive (Li & Sung, 2021). Specifically, humanizing AI chatbot reduces customers' perceived risk, enhances customers' trust and credibility in AI chatbot (Murtarelli et al., 2021). In addition, interacting with AI chatbots satisfies the social desires of customers who have

high social needs (Sheehan et al., 2020). As NLP technology matures, human-like features of AI chatbots will become more important for customers to obtain services from AI chatbots (Sheehan et al., 2020).

In the text data, we find many descriptions related to human-like. Both customers and organizations believe that human-like characteristics are important aspects of the AICSQ. Customers perceive the human-like characteristics of chatbots through social cues such as voice, image, name, and human-like personality. In addition, human-like empathy is a factor that customers attach much importance to. From the dialogue with AI chatbots, customers perceive that they are cared for and valued. High empathy is regarded as a feature of high service quality (Parasuraman et al., 1988). The text examples are shown in Table 2.

Continuous Improvement: One of the significant differences between AI and human customer service is that machine learning speed is much higher than that of human learning. After a period of use, customers hope that the AI chatbot can continuously update its Q&A database, answer more questions, understand the problems more accurately, and solve problems better. These enhancing abilities are the results of self-learning. Both customers and organizations hope for AI chatbots to improve through self-learning continuously. The organization's interview data also show that a system update is a factor in evaluating AICSQ. For example, after the update, the system is more stable, and the database is more comprehensive. Table 2 gives the text examples.

Personalization: Many complaints about AI chatbots come from the AI chatbot services that are not personalized enough. AI chatbot is professional and efficient in dealing with standardized and routine questions. However, AI chatbots often cannot handle highly complex and rare requirements,

negatively affecting users' perception of AICSQ. Big data technology makes it possible for AI to provide personalized services. In the text data, we found that the primary factor of personalization is that AI chatbots can identify a customer's unique characteristics based on the customer's big data, such as demographic information, hobbies, personality, and political stance, so as to respond appropriately to the customer. Customers hope that AI chatbots can meet their specific needs and give them personalized responses. In addition, a personalized recommendation based on big data is an advantage of AI chatbots compared to human customer service (Wirtz et al., 2018). The accuracy of the personalized recommendation of AI chatbots also reflects its service quality.

Culture adaption: Most of the 55 organizations we interviewed have overseas operations, and these organizations with overseas operations address the importance of AI chatbots' culture adaptability, which mainly includes two aspects: First, AI chatbots need to have multi-language ability to communicate with customers from various countries. The interviews data indicates that customers hope that AI chatbots can understand their native language when they do shopping on a foreign E-commerce website. Second, AI chatbots need to understand the cultural differences of countries and adapt to customers' collective psychological characteristics in different countries. For example, the text data in Table 2 shows that customers with different cultures have distinct attitudes and expectations towards AI chatbots and represent different acceptance levels of standardized or personalized services.

Efficiency: Under this category, we identify the factors of efficiency, which reflect the high quality of AI chatbot service. This dimension includes three aspects: G1. Always available. AI chatbots are always available 24 hours×7 days. AI chatbots are essentially software embedded into computers or mobile devices. Customers can find AI chatbots and use them anytime and anywhere. G2.

Responsiveness. One of the aspects the users satisfy with AI chatbots most is responsiveness. Anytime customers make requests to AI chatbots, they receive feedback immediately. G3. Quicker service. The service provided by AI chatbots is not only to answer customers' questions but also to simplify many service processes by providing quick services. For example, AI chatbot can send logistics information, product choice link, or payment link to customers so that they can operate on the interface directly and conveniently. The automated service allows customers to enjoy efficient service.

—————————————————

Insert Table 2 here

—————————————————

The above seven selective codes and 18 axial codes are the second-order and first-order dimensions of the AICSQ, respectively. The definitions of the constructs are shown in Table 3 in Section 3.2.

3.2 Phase 2: The scale development of AISCQ

This section aims to develop the scale of AICSQ. We referred to Mackenzie et al. (2011)'s ten-step method to develop the scale. Phase 2 mainly includes item generation and scale purification.

This section presents the first four steps in the ten-step method. Step 1: Define all the constructs. We defined all the dimensional constructs which we obtained from text coding in the last section. We defined the constructs based on the text data and related literature (shown in Table 3).

—————————————————

Insert Table 3 here

—————————————————

Step 2: Generation of items. Some of the constructs were new with no instruments in the existing literature, so that we developed the item pool from the open codes. Some constructs had measurement instruments. However, the existing instruments could not reflect the AI context. So we adaptively

revised the existing scale based on AI chatbot service context. The initial scale with 111 items was formed. Step 3: Content validity test. We recruited two Ph.D. students in information systems to conduct a content validity test. Following Anderson & Gerbly (1991), we first explained all the definitions of constructs, showed all the items to them, and then asked them to place the items under the constructs, which could reflect the meaning of the constructs best. After they finished the work, we discussed the inconsistency and revised the vague items. Step 4: Develop a measurement model. We constructed a measurement model with seven second-order- and 18 first-order- constructs. For each second-order construct, the sub-dimensional first-order constructs reflected an aspect of them, and the measurement model was reflective-formative type. The measurement model is shown in Figure 3.

————————————————————
Insert Figure 3 here
————————————————————

Then we purified the scale to develop a scale with reliability and validity. We collected two rounds of survey data in step 5 and step 7. In step 5, we collected the first round of data from the survey data collection platform in China, www.wjx.com, which was the biggest market survey company in China. The questionnaire in this round was in Chinese. We showed the English and Chinese versions in Table 2 in the supplementary material (the process of three-round survey data examination of the scale development). We used the backward translation method to ensure translation validity. We recruited two Ph.D. students who were familiar with the research topic to conduct the translation. The scale was first in Chinese and one Ph.D. student translated it into English. Another Ph.D. student translated it back into Chinese. Then we compared two versions and found there were no differences. To ensure the qualification of the respondents, we adopted two methods: on the one hand, we asked the platform to screen the respondents for us, requiring the respondents must have used any type of AI chatbot

16

service. We paid an extra 10 RMB for each questionnaire for the screening service. On the other hand, we set a question in the questionnaire that asked whether the respondent used any type of AI chatbot and which organization's AI chatbot had they used. We paid twenty-five RMB for each questionnaire in total. When we sent out the electronic questionnaire, we briefly introduced the purpose of the survey: "This questionnaire aims to investigate customers' evaluation of AI chatbot service quality. Filling this questionnaire will take you 10 to 15 minutes. If you have never used an AI chatbot, please do not fill in the questionnaire. Please recall your experience of interaction with the AI chatbot and fill in the questionnaire based on your true feelings. All responses are for academic research purposes only." Finally, we collected 490 questionnaires in round-1 data collection, and after deleting invalid responses, 442 valid questionnaires were left. We relied on three standards to evaluate whether the questionnaire was valid or not. First, we deleted the respondents that filling time was below eight minutes, which was the minimum time tested by the authors. Second, we found the contradiction based on the answers to the reverse question. Third, all the answers were the same except for the questions of demographics. 52.2% of the respondents were male and 47.8% were female. 39% were between 31 to 40 years old, and 30.7% were between 26 to 30 years old.

Step 6 is the scale purification. We used SmartPLS 3.2.4 to conduct the measurement model testing. SmartPLS is suitable for the explorative study in which the relationships between variables are new and untested (Gefen & Straub, 2005). Moreover, SmartPLS has advantages in testing models that contain formative constructs, which are the second-order constructs in our measurement model. We first used SPSS to conduct an exploratory factor analysis (EFA) test. We extracted 18 factors with eigenvalue above one, which explained 75.6% of the total variance. We adopted the maximum variance

method, and based on the result of varimax rotation factor loadings, we deleted the weak items whose factor loadings are below 0.7. After deleting the weak items, we recalculated the varimax rotation factor loadings. All the factor loadings are above 0.7, and the loading values of items on the corresponding constructs are far more than that on other constructs. Then we examined the reliability and validity. Testing results show that the value of Cronbach' s alpha and CR values were above 0.7, which indicates the scale had good reliability (Nunnally, 1978). The values of AVE were above 0.5, indicating good convergent validity (Bagozzi & Yi, 1998). After comparing the square roots of the AVEs and correlations between constructs, we found the former is larger than the latter, indicating that the discriminant validity was acceptable. After scale purification in this round, we deleted ten items and 101 items left.

Following step 7, we collected the second-round survey data on the platform zbj.com. We paid each respondent 10 RMB. After two weeks, we obtained 521 questionnaires, and 471 were reserved after deleting the invalid ones. The demographic of respondents shows that 50.3% were male and 49.7% were female. 39.6% were between 31 and 40 years old, the most significant proportion of the age composition. Step 8 is the scale repurification. We first repeated the scale refinement steps in the last round of data testing. Based on the results of the factor loading, we deleted seven weak items. Then we conducted the EFA test again, and the results of factor loading were acceptable. We examined the reliability, discriminant validity, and convergent validity by testing the value of Cronbach' s alpha, CR, AVE, and compared the square roots of the AVEs and correlations between constructs. The above results suggest that reliability and validity were passed. Besides, to test the relationship between first- and second-order constructs, we tested the VIF value of the first-order constructs, and they were below

the threshold of 3.3 (Diamantopoulos & Siguaw, 2006). The results of the relationship between first-order and second-order dimensions show that the first-order constructs significantly contribute to the second-order constructs. Further, we examined the possible CMB by setting up a latent method factor in the model and calculated the average quadratic sum of the principal variable loadings and the method factor loadings (Liang et al., 2007). The results show that CMB was not a serious problem in the two rounds of data collection. The results of the CMB test are shown in Table 12 and Table 13 in the supplementary material of "the process of three-round survey data examination of the scale development". After the two-round scale refinement, we obtained the AICSQ scale with 95 items. The refined scale is shown in Table 4.

—————————————————

Insert Table 4 here

—————————————————

3.3 Phase 3: Developing norms

In phase three, based on the framework of quality–value–satisfaction–loyalty chain(Parasuraman & Grewal, 2000) and IS success model (DeLone & McLean, 1992), we constructed an AI chatbot service success model to conduct a nomological test, which is shown in Figure 4.

—————————————————

Insert Figure 4 here

—————————————————

Previous studies empirically demonstrate the validity of the quality–value–satisfaction–loyalty chain(Cronin et al., 2000; Parasuraman & Grewal, 2000). This framework can be well integrated with the IS success model to study the success of e-commerce systems (Wang, 2008). DeLone & McLean (1992) propose an IS success model after they comprehensively review the studies on evaluation of IS

success. The initial IS success model contains six factors: system quality, information quality, IS use, user satisfaction, individual impact, and organizational impact. As the development of e-commerce, many researchers investigate the factors of e-commerce system success (Liu & Arnett, 2000; Molla & Licker, 2001; Palmer, 2002). In this background, DeLone & McLean (2003) update the IS success model and add new factors of service quality, intention to use, and change the consequence factor into net benefit. Although the updated model is generic, it is inconsistent with the nomological structure in marketing literature. To reconcile this inconsistency, Wang (2008) respecifies the IS success model by integrating the quality-value-satisfaction-loyalty chain. In the respecified model, e-commerce system success factors include information quality, system quality, service quality, perceived value, user satisfaction, and intention to use. Intention to use represents customers' loyalty, which is a net benefit of an organization.

The research context of this study is customer-AI chatbot interaction in e-commerce. Essentially an AI chatbot is an advanced system. Therefore, we use the respecified IS success model to conduct the nomological test (predictive validity test). AI chatbot is different from the traditional information system. AI chatbot is not only a type of information system but also a service provider in the service frontline. AI chatbots provide service through providing information. Thus, in AI chatbot service context, system quality, service quality, and information quality are inseparable. In addition, as the AI chatbot is distinct from the traditional information system, the service quality of AI chatbots needs to be reconsidered. In Section 3.1, we coded seven second-order dimensions and 18 first-order dissensions based on interview text to evaluate AI chatbot service quality.

In this AI chatbot service success model, the critical factors of AI service success include service

quality, the perceived value of AI chatbot, satisfaction with AI chatbot service, and customers' intention of continuous use. Compared with the system in the context of DeLone & McLean (1992), AI chatbot has some intelligence of human beings and can replace humans to interact with customers to some extent. By constructing the dimension of AICSQ, we have opened the black box of service quality in the AI service context.

Customers' perceived value comes from evaluating the relative rewards and sacrifices associated with the product or service (Yang & Peterson, 2004). In the AI chatbot service context, if customers perceive the time and effort they put in exchange for high-quality service as valuable, then they would perceive the high value of the chatbot. Prior studies have demonstrated the positive relationship between service quality and perceived value in the service context (Cronin et al., 2000; Parasuraman & Grewal, 2000). AI chatbot service quality contains seven dimensions: semantic understanding, close human-AI collaboration, human-like, improvement, personalization, cultural adaption, and efficiency. Thus, we hypothesize that:

H1a to H7a: The AICSQ of semantic understanding, close human-AI collaboration, human-like, continuous improvement, personalization, culture adaption, and efficiency are positively associated with customers' perceived value of AI chatbot.

In the service process, customer satisfaction is an overall subjective evaluation of the service or transaction process, and it is a rising emotional state(Oliver, 1981).When customers perceive high service quality, they perceive that AI chatbot service satisfies the certain needs of them or the service quality they provided surpassed their expectations, making them satisfied with the service provider (Harris & Goode, 2004; Maklan et al., 2017). Regarding the relationship between perceived value and

satisfaction, perceived value is the result of the cognition-oriented evaluation, and satisfaction is an emotion-oriented psychological response. Cognition-oriented value evaluation precedes emotional response(Gotlieb et al., 1994). In addition, the existing literature validates the value-satisfaction-loyalty chain and uses this VSL framework as a theoretical basis to analyze customer loyalty (Xu et al., 2015; Yang & Peterson, 2004), which demonstrates the positive relationship between perceived value and satisfaction. In this research context, when customers perceive an AI chatbot as with high service quality and high value, they would be more satisfied with the AI chatbot. Hence, we hypothesize that:

H1b to H7b: The AICSQ of semantic understanding, close human-AI collaboration, human-like, continuous improvement, personalization, culture adaption, and efficiency are positively associated with satisfaction with AI chatbot service.

H8: Customers' perceived value of AI chatbot service is positively related to satisfaction with AI chatbot service.

From the IS success model perspective, customer satisfaction is positively related to the net benefit. Customers' willingness to continuously use is a typical net benefit for organizations. Also, an obvious representation of customer loyalty is to continue to use the organizations' service or products, reflecting customers' action loyalty (Harris & Goode, 2004). If customers perceive that AI chatbot service is valuable to them, they would have high chance to obtain service again (Wang, 2008). In addition, based on the framework of service quality-value-satisfaction-loyalty chain, we hypothesize that:

H9: Customers' perceived value of AI chatbot is positively associated with intention of continuous use.

H10: Customers' satisfaction with AI chatbot service is positively associated with intention of

continuous use.

To examine the measurement model and structure model, we collected the third round of survey data. The scale of perceived value was adapted from Bernardo et al. (2012), satisfaction was adapted from Fang et al. (2014), and continuous use was adapted from Bhattacherjee (2001). The scale is shown in Table 14 in the supplementary material of "the process of three-round survey data examination of the scale development". In this round of data collection, we used international samples to test the measurement and structure model. We designed the questionnaire in Qualtircs and distributed them through a panel of Prolific. From the demographic information that Prolific provides, the respondents were from the United Kingdom, Australia, France, Italy, Spain, Canada, Israel, Greece, Belgium, Poland, Mexico, South Africa, Slovenia, Estonia. We collected 663 samples and 597 valid respondents left after screening. Regarding the composition of respondents, 47.7% of them were male and 52.3% were female, and 45.2% of them were between 18 to 24 years old and 37.5% of them were between 25 to 34 years old. We first conducted a measurement model test using the previous method, which is the Step 9 of cross-validation. The results of varimax rotation factor loadings, the square roots of the AVEs and correlations between constructs, the values of Cronbach's alpha, CR, and the AVE values all indicate good reliability, convergent validity, and discriminant validity of the scale. Besides, we examined the relationships between all the first- and second-order constructs. We presented the results of the path coefficient and T-value of the first- and corresponding second-order constructs and the VIF value of all the first-order constructs. The results demonstrated the formative attribution of the second-order constructs. The CMB test results show that the CMB was not a serious problem in this dataset. The above results are shown in Tables 15 to 20 in the supplementary materials. In Step 10, developing

the norm helped explain the research findings and guided future research. The SEM test results are shown in Table 5. The R^2 of satisfaction and intention of continuous use were 0.34 and 0.28, respectively, indicating the total variances explained were 34% for satisfaction and 28% for continuous use. The path coefficients support our hypothesis 1a-7a and 1b-7a, i.e. all the AICSQ dimensions are positively associated with customers' perceived value of AI chatbot and their satisfaction with AI chatbot service ($p < 0.05$ for all path coefficients), showing the good nomological validity of AICSQ scale. Moreover, the SEM model supports our hypothesis 7-10 ($p < 0.01$), indicating that the customers' perceived value of AI chatbot and satisfaction with AI chatbot service are positively associated with their intention of continuous use of the service.

—————————————————————

Insert Table 5 here

—————————————————————

4. Discussion

As AI chatbots enter the frontline to service customers and service quality is vital to organizations, the existing dimensions and scales of service quality are challenged. This study establishes the dimension system of AICSQ based on clarifying the unique characteristics of AI chatbot service and develops the associated scale. We adopt an integrated perspective of organizations and users to conduct this study, avoiding the bias under a single view. This dual perspective is mainly reflected in the analysis and verification of data. The textual data for dimension construction was mainly collected from the interviews of 55 organizations from 17 countries, while the data for scale development, verification, and the nomological test was collected from users. This integrative method analysis with multiple sources of data helps us understand AICSQ by standing on both customers' and organizations'

positions. Specifically, in phase 1, we developed open codes of the semi-structured interview text and gradually clustered them to axial codes and selective codes. Axial codes and selective codes together form a two-dimensional system of AICSQ , which is shown in Figure 3. The second-order dimensions of AICSQ include semantic understanding, close human-AI collaboration, human-like, continuous improvement, personalization, culture adaption, and efficiency. In phase 2 and 3 we adopted the 10-step method to develop the scale for these dimensional constructs in three phases. After checking the content validity of the initial scale, we examined the factor loadings, reliability, and validity of the 3rd round international survey data, and refined the scale accordingly. Finally, we obtained the AICSQ scale with good reliability and validity. Following this, we built an AI chatbot service success model and conducted a nomological test.

4.1 Implications

4.1.1 Theoretical implications

This study contributes to service quality studies in several ways. First, this study fills the gap in the lack of the classification of AICSQ. Service quality is a classic but hot research topic in the IS and marketing field, which has a significant impact on user satisfaction and loyalty. With the advent of large-scale AI chatbots in the market, studies on AI chatbot services have increasingly grown (Luo et al., 2019; Marinova et al., 2017; Ostrom et al., 2019). Supported by AI technology, the service provided by AI chatbot is different from SST and human customer service. The existing dimensions of service quality can hardly be applied to AI chatbot services. The dimensions and scales of AICSQ are the basis for the empirical study on AI chatbot services. The innovative classification of AICSQ reflects the unique technical characteristics of AI. For example, AI chatbots' accuracy of response and

understanding emotion depends on the maturity of NLP technology. The constructs of identify customers and personalized recommendations reflect the AI chatbot machine learning and big data analysis capabilities. Besides, customers' perception of human-like intelligence lies in AI's sense, recognition, analysis, and decision-making abilities. The research on the dimension of AICSQ is blank and the existing dimension of service quality of E-service, websites, and information systems cannot reflect the technical characteristics of AI. The vast majority of dimensions we developed in this study are new, which are generated from the interview data analysis, not from the existing literature. The new constructs of AICSQ dimension include semantic understanding, close human-AI collaboration, human-like, culture adaption, and efficiency in second-order constructs, and understanding query，understanding emotional expression, easy to transfer, human-like social cue, human-like personality, human-like empathy, self-learning, identify customers, personalized response, non-language barriers, culture understanding, and process simplify in the first-order constructs. Regarding the small minority constructs which already exist in the literature, such as accessibility, memorability, personalized recommendation, we adapt the existing scale by considering the new AI context and interview data. Therefore, the multi-level dimensions of AISCQ capture the AI technical features, which extend the service quality research in the AI context.

Second, the scale of AISCQ we developed provides measurement instruments for subsequent scholars to conduct empirical research on user behavior when interacting with AI chatbots. Some of the service quality constructs in our multi-dimensional system have existed in previous literature, and some are new constructs. For the former, we find that the existing scales could not be adapted to the new AI context directly. Based on the open codes and existing scales, we rewrote the items of these constructs. For the

latter, we wrote the initial scale based on the definition of the new construction and the open codes in the text data. By adopting the 10-steps scale developing method (Mackenzie et al., 2011), we examined and refined the scale by testing three rounds of data and finally obtained AICSQ scale with good reliability and validity.

Third, we developed the IS success model from two aspects. On the one hand, we developed the critical factors of AI chatbot service success based on the context, including AICSQ, customers' perceived value of, satisfaction with, and intention to continuous use of AI chatbot service. Compared with DeLone & McLean's IS success model, we opened the black box of AICSQ and examined the relationship between each service quality dimension and other service success factors. In the original IS success model, service quality is divided into information quality and system quality. In the context of AI, we find that in AI chatbot service context, the system quality, information quality, and service quality are inseparable. AI chatbot is a kind of information system, and it serves customers by providing information. In addition, based on the definition of service quality, which is a perception and is the overall subjective evaluation of the service (Dabholkar et al., 2000; Parasuraman et al., 1988), AICSQ contains users' perception of system and information quality. This study develops an integrated evaluation system of AICSQ, which reflects the success of AI chatbot service.

4.1.2 Practical implications

Organizations use AI chatbots mainly in three ways: developing AI chatbots by themselves, purchasing AI chatbots from information technology firms that specializes in producing AI chatbots, and directly using the chatbot on social media platforms, such as Expedia Facebook Messenger Bot. Thus, there are three types of stakeholders for AI chatbots: 1) AI chatbot developers, including the organizations'

internal development department and external development firms, 2) the organizations that employ AI chatbots, and 3) AI chatbot end-users or end-customers. Our research result contributes to all three types of stakeholders.

First, for the developers of AI chatbots, the multi-level dimensions and scale of AICSQ play an important role in the entire AI chatbot development cycle. During the investigation phase, the dimensions and scales of AICSQ can be used as a reference for feasibility analysis. Developers can understand the most critical factors of AI service quality from this dimension and check whether their resources are enough to create an AI chatbot with high service quality or not. In the development stage, AICSQ dimensions can be used as a checklist in the design and production process. Finally, in the market launch stage, developers can use them to evaluate the service quality level of an AI chatbot.

Second, for AI chatbot employers, organizations hope to partially replace human customer service with AI chatbot and achieve certain business goals such as simplifying service processes, increasing service efficiency, customer satisfaction, and profits. The foundation of achieving these business goals is to guarantee high AICSQ. Our research results provide organizations with a valid reference for choosing AI chatbots. Before purchasing or "hiring" an AI chatbot, organizations can evaluate whether the AI chatbot will be a competent employee or not based on our dimensions. In addition, directly using the AICSQ scale to investigate the satisfaction level of end-customers on the AI chatbot and collecting feedback for the subsequent AI chatbot improvement would contribute to the companies' service quality control and performance.

Third, for the end-customers, they can learn about the advantages of AI chatbots through our dimensions and scales to explore and make sense of the AI chatbots. In the era of service-dominat

logic, customer participation in value co-creation is an important feature of business(Barrett et al., 2015; Lusch & Vargo, 2014). The proposed AICSQ dimensions make customers more "professional" in experiencing AI supported service, and in turn, benefit the organizations in improving service quality based on these "professional" customer feedback. When customers participate in co-creation activities such as the "Customer Experience Improvement Program", they can also use the AICSQ dimensions and scales of this research to propose valuable suggestions. The suggestions directly hit the core of AI service quality are easier to be adopted.

4.2 Limitation and future research

In addition to extending previous research in service quality in the digitalized business, we identified and measured the construct of AICSQ using samples of diverse consumers from multiple countries and sectors. One of the limitations of this study is the incomplete research on the types of AI chatbots. According to ISO 8373:2012, service robot refers to robots that perform useful tasks for humans or equipment, excluding industrial automation applications. Future research can refine the conceptualization and measurements in specific contexts with other types of AI chatbots. The AI chatbots in the market include online virtual types as well as offline physical AI chatbots, such as human-shaped customer service robots in the lobby of a bank or hospital. In different contexts, service robots execute different tasks. Therefore, customers have different requirements and expectations to them. Most consumers have experience in interacting with AI chatbots in e-commerce. The dimensions and scale developed in this study reflect the features of AI chatbots in an online e-commerce context. In other contexts, service robots may be designed to achieve specific aims. Therefore, the service quality dimensions may contain context-specific factors. For example, for healthcare service robots,

the customers care more about the service quality related to health care functions. In this case, future research can explore the dimensions and influential factors of service quality of service robots in the specific task context.

The impact of human-likeness of AI chatbots needs to be further explored. In this research，we coded human-like as a dimension of AI chatbot service quality based on qualitative data. We find both customers and companies attach importance to the human-like features of AI chatbots. They expect that AI chatbots "provide warm service, are friendly, have the ability to be approachable, and are good listeners". However, Lance & Jill (2019) suggest that even under anthropomorphism, customers can clearly realize that an AI chatbot is basically a programmed robot. Prior study also finds that in some specific contexts, anthropomorphism may not be necessarily good. For example, when a customer is angry, anthropomorphism may decrease their satisfaction (Crolic et al., 2021). In the future study, it is interesting to explore the effect and internal mechanisms of human-likeness of AI chatbots in specific contexts.

Our analysis highlights the need to investigate further how the evaluation on the service quality of AI chatbot would affect the customers' continuous use of intention. The perceived value of AI chatbots and the satisfaction with AI chatbots are two key mediators in the relationship between AI chatbot service quality and customers' continuous use of intention. One direction for future research is to consider how these two factors would affect customers' continuous use of intention in different industries and with different types of AI chatbots. For example, the perceived value of an AI chatbot may be more important for users in a knowledge-intensive organization, while the satisfaction with an AI chatbot is more important for the users in an online-shopping context. Moreover, future studies may

include other factors, such as the trust of AI chatbots, which affect the customers' continuous use of intention.

5. Conclusion

This study established the multi-level dimensions of AICSQ and developed the associate scales by adopting the mixed-method approach. Based on the interview data from both organizations and customers, we conducted open coding, axial coding, and selective coding. Finally, we obtained a seven second-order-dimension and 18 first-order sub-dimension AICSQ classification system. The ten-step method was adopted to develop the measurement scale. After obtaining a reliability and validity scale, we built and examined the AI chatbot service success model. The study contributes to academics by providing an innovative dimension and scale system of service quality in the AI context and contributes to practice by giving three kinds of stakeholders specific suggestions based on our research results.

Tables and Figures

Table 1 The difference in service quality between AI, human customer service, and SST

Comparing types	The distinction of service quality between AI chatbot and human customer service		The distinction of service quality between AI chatbot and SST	
	AI chatbot	**Manual human customer service**	**AI chatbot**	**SST**
Service methods and improvements	Improve service quality through self-learning; The learning source is big massive data	Improve service quality mainly through training; The learning sources and input are limited	The interaction process is flexible; AI chatbot guides customers in a human-like manner	Users need to learn how to use SST; Need to comply with SST processes and rules
Service outputs	Homogeneous output; The service quality is consistent and stable	Heterogeneous output; The service quality is unstable, which depends on the state and emotion of the service provider	AI chatbot can tolerate mistakes. That is, even there are certain mistakes in the user's operation, the process can also be executed	If SST is used incorrectly, the function cannot be executed
Service functions	They can achieve the personalized and accurate recommendation derived from the big data analysis; They can achieve a certain degree of personalized interaction	They can achieve a certain degree of accurate recommendation; They can achieve a high degree of personalized interaction	Like human customer service, AI chatbot can guide customers when they conduct wrong operations	Difficult to recover automatically after service failures
Emotions in the service process	They can simulate emotions and can affectively interact with customers shallowly	They can produce emotions and interact with customers deeply emotionally	It has the task-oriented practical value and hedonic value	Focus on the task-oriented practical value

Table 2 The examples of text data, open codes, axial codes, and selective codes

Text data example A1-A2

A1. While response of AI-based chatbots are usually more prompt than manual service, accuracy of response must be ensured. It is often this aspect that might result in a bad user experience despite providing timely service results...It's not smart enough, and it cannot answer complex requests... Even though AI-based chatbots are programmed to understand the queries of users that are typed in natural language, they might misunderstand and

often ask users to repeat the question.

A2. Computer-human interaction should be able to identify the current emotion status of customers during chatting, and there use difficult attitudes and emotions to treat customers...AI-based chatbot service should chat with emotion

Open codes	Axial codes	Selective code
Accuracy of response, Understand complex question	A1. Understanding query	A. Semantic understanding
Identify the current emotional status, use difficult attitude and emotion to treat customer, understand emotion	A2. Understanding emotional expression	

Text data example B1-B4

B1. The human customer service receives a message from the bot when it cannot answer. The customer is notified that the human customer service is going to take over the issue in a certain time...The intelligent customer service is still for assistance and simply tasks. However, human customer service deals with more complex and specific issues...We have an intelligent information systems to transfer jobs from intelligent customer service to the human customer service, and also in another direction...If the query is complex and is not resolved using the AI, then it gets routed to human customer services where live agent can solve the problem. AI usually has common queries encoded in it, so a coordination between AI and human customer service is required...Compared with human customer service, easy to contact of AI-based chatbot should be considered seriously

B2. When AI chatbot cannot solve the requests, customers can find human customer service easily.

B3. It's not smart. It's only well engineered for answering to simple and regular requests. When the question is about something it can't solve, a customer service takes the conversation over. Customers need not to repeat their requests...Remember previous communications should be considered seriously.

Open codes	Axial codes	Selective code
Easy to contact, transfer jobs, customer service takes the conversation over	B1.Accessibility	B. Close Human-AI Collaboration
Transferring to or cooperating with human customer services easily.	B2.Easy to transfer	
need not to repeat requests when transfer service	B3.Memorability	

Text example C1-C3

C1. The designed interactive language and smiling face of AI reflects the intelligence...Ann's picture, voice and natural language interpretation are aspects to be considered...users can talk directly to "Julie," Amtrak's customer service bot.

C2. Historically, humans have been providing services. Thus, to maintain the level of comfort that users are already accustomed to, many AI-based chatbots are designed to display humanoid personality.

C3. Computer-human interaction should be friendly... Ability to be approachable. Bots can be a good listener.

Open codes	Axial codes	Selective code
Interactive language, smiling face, name	C1.Human-like social cue	C. Human-like
Humanoid personality	C2.Human-like personality	

Courtesy, friendly, approachable, good listener	C3.Human-like empathy

Text example D1-D2

D1. Reinforcement learning: like if something is not coded in a system and it is not able to solve a query, it should be able to solve it next time...Collect the information from internet or big data and match the requirement of user. The chatbot should be smart to learn the behave of particular user such as what he/ she interests in.

D2. We update techniques faster...AI is always improving, accumulating requests solved in its database.

Open codes	Axial codes	Selective code
Reinforcement learning, learn the behave of particular user	D1.Self-learning	D. Continuous Improvement
Update techniques	D2.System update	

Text example E1-E3

E1. Computer-human interaction should be able to identify the sex, the age of customers, and they use difficult attitude and emotion to treat customers. For example, children and elderly persons expected different attitudes... should be able to identify the education level of customers, the religion of customers, the political positions of customers, the personality of customers, the hobby preferences of customers, the wealth status of customers, the profession of customers, and there use difficult attitude and emotion to treat customers.

E2. Services tailored to meet the specific needs of the customer, rather than offering a platter of general choices, will enhance user experience, in a manual service system, this becomes an in-built feature whereby the service provider responds to the specific query and two-way interaction happens...Chatbot should be able to answer the query depending upon the past purchase behavior of the customer ...People complain that functionality is limited and it is not personalized.

E3. We believe that AI in our situations provides improve quality in the personalized recommendation. Some Feel Life features that AI-based chatbots have are: Humanoid personality, personalized Recommendation and Customization.

E4. AI-based Chatbot may be tedious during chatting or the interaction should not be only on side. Chatbot should also be active to enquiry the information needed from customers based on the detail pictures of customers.

Open codes	Axial codes	Selective code
Identify the sex, identify the age of customers, identify the education level, identify the religion of customers, identify the religion of customers, identify the political positions of customers, identify the personality of customers, identify the hobby preferences of customers, identify the profession of customers	E1.Identify customers	E.Personalization
Use difficult attitude and emotion to treat customers, meet the specific needs of the customer	E2.Personalized response	
Personalized Recommendation, Customization.	E3.Personalized recommendation	

Text example F1-F2

F1. AI can use all the languages that facebook uses (English, French, Spanish, German and Russian)

F2. Culture understanding is important for computer-human interaction. For example, a French and a German may have different feelings in chating with bots because culture difference. Different in the dimension of "Individualism vs Collectivism". A Chinese and a French may have different attitudes to chatting with robots.

A Chinese with higher Collectivism may accept some standard service. But a French with higher Individualism may expect more customized service...National culture difference. For example, a French and American may have different expectations in communication with chatbots....culture adaption with different peoples of AI-based Chatbot should be considered seriously.

Open codes	Axial codes	Selective code
Use all the languages	F1.Non-language barriers	F.Culture adaption
Culture understanding, culture adaption with different peoples	F2.Culture understanding	

Text example G1-G3

G1. Aetna's secure member website. "Ann," Aetna's personalized, virtual assistant, offers 24-hour support for members who are new to the website or need help logging in. ... Eva is accessible through mobile, internet and sms banking services... AI-based chatbot is also omnipresent.

G2. Compared with human customer service, speed of response of AI-based chatbot should be considered seriously... it should look at the history of the customer purchase behavior to understand the questions properly. It should be Interactive, User interface should be pleasing and questions should be answered instantaneously

G3. The AI-based chatbot can recognize the customer's demand and reply to them immediately without any waiting. Our company believes that using the AI-based chatbot is a faster way for our customers to deal with their issues, so that our customers can save more time and increase the satisfaction...The AI is really fast and can make an order faster than the human customer service if the human being interacting with the chatbot is precise in his/hers writing. ... I like to use AI chatbot to charge my telephone fees. After I type my phone number, the chatbot automatically send me different amounts to choose and send the payment link. It is really convincing and efficient.

Open codes	Axial codes	Selective codes
Bot service is always available, offers 24-hour support for members, works 24/7,available at all times, mobile, omnipresent	G1.Always available	G. Efficiency
Quick responses, Speed of response	G2.Responsiveness	
Make an order faster, quicker service , automatic	G3.Simplify process	

Table 3 The definitions of constructs

Constructs	Definition	Source of definition
A. Semantic understanding	AI chatbots can understand the meaning of the content sent by the customer in the form of voice or text.	(D'Mello et al., 2010) and open codes
A1. Understanding query	AI chatbots can understand customers' query well.	open codes
A2.Understanding emotional expression	AI chatbots can understand the customers' emotional state through the interaction with them.	open codes
B.Close Human-AI Collaboration	AI and manual customer service provide services to customers together, complement each other's advantages, and expect to achieve a common goal.	(Donmez et al., 2009; You & Robert, 2018) and open codes

B1.Accessibility	Customers can easily obtain AI and manual customer service in the interface from the system.	(Wixom & Todd, 2005)
B2.Easy to transfer	When AI or manual customer service cannot solve a customer's request or meet customer's needs, customers can easily switch from manual customer service to AI chatbot, or in another direction.	Open codes
B3.Memorability	The system can remember the customer's questions, needs or preferences, so that when the task is switched from AI chatbots to manual customer service (or vice versa), customers do not need to repeat their requests.	(Lacka & Chong, 2016) and open codes
C.Human-like	AI chatbots have human-like qualities and interacts with the customers in a natural, human way.	(Rijsdijk et al., 2007; Schroll et al., 2018) and open codes
C1.Human-like social cue	People can perceive AI chatbots' human-like social cues, such as human-like voice, avatar, name.	(Qiu & Benbasat, 2009) and open codes
C2.Human-like personality	Customers can perceive AI chatbots' human-like personality when interacting with them.	(Mourey et al., 2017)
C3.Human-like empathy	Customers can perceive AI chatbots' caring, individualized attention, and understanding when interacting with them.	(Parasuraman et al., 1988) and open codes
D. Continuous Improvement	The continuous improvement of AI chatbots capabilities.	(Obert et al., 2000) and open codes
D1.Self-learning	AI chatbots can conduct continuous self-learning and improve the ability to solve problems based on big data analysis.	(Rijsdijk et al., 2007) and open codes
D2.System update	The continuous maintenance, repair, upgrade and update of AI chatbots software.	(Overgoor et al., 2019) and open codes
E. Personalization	How well the customer feels that the AI chatbots's service can be personalized to meet his or her needs.	(Kim, 2009) and open codes
E1.Identify customers	AI chatbots can identify customer portraits based on customer big data.	Open codes
E2.Personalized response	The AI chatbots can generate appropriate response using the user input features.	(D. Lee et al., 2017) and open codes
E3.Personalized recommendation	AI chatbots can make personalized recommendations which reflects customer needs and preferences.	(Tarn & Ho, 2005) and open codes
F. Culture adaption	The ability of AI chatbots to adapt to the culture of host countries and to Interact with citizens of different cultures.	Open codes
F1.Non-language barriers	AI chatbots can freely use various languages to communicate with customers from different countries without barriers.	Open codes
F2.Culture understanding	AI chatbots can understand the culture of different countries well, so as to respond appropriately to customers from different countries.	(Watkins & Gnoth, 2011)
G. Efficiency	AI chatbots can serve customers efficiently.	Open codes

G1.Always available	AI chatbot is available 7 days × 24 hours. Customers can get AI chatbots at anytime and anywhere.	Open codes
G2.Responsiveness	Customer perceived that AI chatbot's willingness to help customers and provide prompt service.	(Watson et al., 2006) and open codes
G3.Process simplify	AI chatbot helps simplify the service process through automation and the service efficiency is improved.	open codes

Table 4 The scale of AICSQ (AI chatbot service quality)

Constructs	Items
A1. Understanding query	1. The AI chatbot can accurately understand my query
	2. The response to my query from the AI chatbot is accurate
	3. The answer of the AI chatbot corresponds to the question I asked
	4. The answer from the AI chatbot meets my expectations
	5. AI chatbots can answer my questions accurately
	1.AI 客服能够精确地理解我的需求
	2.AI 客服的回复很准确
	3.AI 客服的回答与我所提的问题是对应的
	4.AI 客服对我需求的回答符合我的期望
	5.AI 客服可以很准确地回答我的问题
A2.Understanding emotion	1. The AI chatbot can recognize my emotions through interaction
	2. The AI chatbot can detect my emotions through interaction
	3. The AI chatbot can perceive my mood
	4. The AI chatbot can make an appropriate reaction to my emotions
	5. The AI chatbot knows if I am happy or unhappy at the moment
	6. The AI chatbot can understand my mood at the moment
	1.在与 AI 客服交流时，他可以识别我的情绪
	2.AI 客服在觉察到交互中我的情绪状态
	3.AI 客服能够感知到我的心情
	4.AI 客服能够对我的情绪做出适当的反应
	5.AI 客服知道我当下是不是开心
	6.AI 客服可以理解我当时的心情
B1.Accessibility	1. I can easily get the AI chatbot or human customer service on the website or APP
	2. It is very convenient and quick to get the AI chatbot or human customer service on the website or APP
	3. I can quickly find the button of AI or human customer service on the interface
	1.我可以很容易地获取 AI 客服或人工客服的服务
	2.获取 AI 客服或人工客服的服务很方便快捷
	3.在页面上可以快速找到 AI 客服或人工客服的服务按钮
B2.Easy to transfer	1. When the AI chatbot cannot solve the problem, I can easily find a human customer service
	2. When the AI chatbot is unable to solve the problem, I can choose a manual service
	3. Manual service and the AI chatbot can easily switch between each other

	4. It's easy to switch from the AI chatbot to a human customer service, or do it in the other direction
	5. I can easily switch between AI and human customer service
	6. Switching service between AI and human customer service is easy for me
	1.当 AI 客服无法解决问题时，我可以很容易找到人工客服
	2.当 AI 客服无法解决问题时，我可以让人工为我服务
	3.人工服务与 AI 服务可以很容易地相互切换
	4.AI 客服转人工服务，或者人工服务转 AI 客服是很容易的
	5.我能够很轻松地在 AI 服务与人工服务之间切换
	6.在 AI 客服和人工客服之间切换对我而言很轻松
B3.Memorabi lity	1. The service system can remember my needs and preferences well
	2. I don't have to submit my requirements to the service system repeatedly
	3. When I switch from the AI chatbot to a human customer service, the human customer service knows my needs
	4. When I haven't used the AI chatbot for a while, it still remembers my preferences or needs after my returning
	5. When I switch from the AI chatbot to a human customer service (or vice versa), the service system will automatically deliver my needs
	1.系统可以很好地记住我的需求、偏好
	2.我不用总是重复提交我的需求
	3.当我从 AI 客服转向人工服务，人工客服知道我的需求
	4.当我一段时间没用 AI 客服，回来后他依然记得我的偏好或需求
	5.当我从 AI 客服转向人工服务（或相反）时，系统会自动传达我的需求
C1. Human-like social cue	1. I can clearly perceive that the AI chatbot has human-like characteristics
	2. I perceive that the avatar or the voice of the AI chatbot is similar to human
	3. The AI chatbot is visually or audibly attractive
	4. The avatar or voice of the AI chatbot is very attractive
	5.The AI chatbot uses the name of human being's
	6. There are no obvious differences between online AI chatbot and human customer service in avatar or voice
	1.我能明显感知到 AI 客服有像人一样的特征
	2.我能感知到 AI 客服的头像形象或声音与人类相似
	3.AI 客服有视觉上或者听觉上的吸引力
	4.AI 客服的头像或者声音很吸引人
	5.AI 客服的名字很像人类的名字
	6.AI 客服在头像或者声音上和人类没有很大区别
C2. Human-like personality	1. The AI chatbot has the similar personality characteristics as humans
	2. I feel that the AI chatbot has its own personality
	3. The personality of the AI chatbot is similar to human being's
	4. The AI chatbot's speaking style is similar to the human being's
	1.AI 客服有像人类一样的个性特点
	2.AI 客服有自己的个性

	3.AI 客服从个性上很像人类
	4.AI 客服说话的语言风格像人类
C3. Human-like empathy	1. The AI chatbot gives me personalized attention
	2. I feel that the AI chatbot puts my interests first
	3. I feel that the AI chatbot serves me attentively
	4. The AI chatbot makes me feel concerned
	5. The AI chatbot makes me feel that it cares about my needs
	6. The AI chatbot makes me feel warm
	1.AI 客服给我个性化的关注
	2.我感觉到 AI 客服是以我的利益至上的
	3.我感觉到 AI 客服在用心为我服务
	4.AI 客服让我有被关注的感觉
	5.AI 客服让我感觉到它在意我的需求
	6.AI 客服让我感觉温暖
D1. Self-learning	1. The AI chatbot can learn from past experience
	2. The AI chatbot can become better through learning
	3. The AI chatbot's ability is enhanced through learning
	4. The AI chatbot can learn to improve themselves well
	5. After a period of use, the AI chatbot's performance is getting better and better
	6. The AI chatbot learns from customers' usage history
	1.AI 客服可以从过去的经验中学习
	2.AI 客服可以通过学习变得更好
	3.AI 客服通过学习能力增强了
	4.AI 客服能很好地学习提升自己
	5.经过一段时间的使用，AI 客服表现得越来越好了
	6.AI 客服可以从顾客的使用历史中学习
D2. System update	1. I can feel the AI chatbot is constantly upgrading
	2. The AI chatbot fixes previous errors
	3. The organizations always maintain their AI chatbot
	4. I feel that the AI chatbot is getting more and more advanced
	5. The function of the AI chatbot has been enhanced
	1.我能感觉到 AI 客服软件在不断升级
	2.AI 客服软件修复了之前的错误
	3.公司在维护他们的 AI 客服软件
	4.我感觉到 AI 客服软件越来越先进
	5.AI 客服软件的功能增强了
E1. Identify customers	1. The AI chatbot can recognize my age
	2. The AI chatbot can recognize my gender
	3. The AI chatbot can recognize my hobbies
	4. The AI chatbot can recognize my buying habits
	5. I feel that the AI chatbot is very familiar with me
	6. I feel that the AI chatbot knows me

	1.AI 客服能识别我的年龄
	2.AI 客服能识别我的性别
	3.AI 客服能识别我的爱好
	4.AI 客服能识别我的购买习惯
	5.我感觉到 AI 客服对我很熟悉
	6.我觉得 AI 客服好像认识我
E2. Personalized response	1. The AI chatbot can give a personalized response to what I say
	2. The AI chatbot can provide targeted answers to my questions
	3. The AI chatbot can make a personalized response to my request
	4. The AI chatbot not always gives a standardized answer
	5. The answer of the AI chatbot is flexible and changeable according to my question
	6. The answer from the AI chatbot make me think it was tailored for me
	1.AI 客服对我说的话能给出个性化的回复
	2.AI 客服能对我提的问题进行有针对性的回答
	3.AI 客服能够对我的要求做出个性化的反应
	4.AI 客服并不总是标准化的回答
	5.AI 客服的回答根据我的问题的不同而灵活多变
	6.AI 客服的回答让我觉得是为我量身定制的
E3. Personalized recommendation	1. I feel that the recommendation by the AI chatbot is in line with my preferences
	2. I feel that the recommendation by the AI chatbot is in line with my taste
	3. The recommendation by the AI chatbot is what I am interested in
	4. The recommendation by the AI chatbot is better than the recommendations I get from other places
	5. I feel that the quality of recommendation by the AI chatbot is what I want
	6. My overall evaluation of the AI chatbot recommendation is very high
	7. I think the AI chatbot recommendations are valuable
	1.我感觉到 AI 客服推荐的内容很符合我的偏好
	2.我感觉到 AI 客服推荐的内容很符合我的品位
	3.AI 客服推荐的内容正是我的兴趣所在
	4.AI 客服推荐的内容比我从其他地方得到的推荐更好
	5.我感觉 AI 客服推荐的质量是我想要的
	6.我对 AI 客服推荐的总体评价很高
	7.我觉得 AI 客服的推荐是有价值的
F1. Non-language barriers	1 The AI chatbot can understand the languages of different countries
	2. The AI chatbot can communicate in various languages without barriers
	3. The AI chatbot can switch languages freely
	4. The AI chatbot can communicate with people from various countries
	5. The AI chatbot can meet my needs for multilingual communication
	6. Even if I use multiple languages to talk to the AI chatbot, it can handle it well
	1.AI 客服能听懂或看懂不同国家的语言
	2.AI 客服可以无障碍地用各种语言进行交流
	3.AI 客服可以自如地切换语言

	4.AI 客服可以与各国的人进行交流
	5.AI 客服可以满足多国语言交流的需要
	6.即使我使用多语言与 AI 客服交谈，他也可以很好地处理
F2. Culture understanding	1. The AI chatbot can understand the culture of my country well
	2. The AI chatbot can understand the culture of my nation well
	3. When I communicate with the AI chatbot, it's like communicating with people from my own country and nation
	4. There is no cultural difference in my communication with the AI chatbot
	5. From a cultural point of view, I communicate smoothly with the AI chatbot
	6. There is no cultural barrier of countries or ethnicities when I communicate with the AI chatbot
	1.AI 客服可以很好地理解我的国家的文化
	2.AI 客服可以很好地理解我的民族的文化
	3.我与 AI 客服交流时，就像与自己国家和民族的人交流一样
	4.我与 AI 客服的交流上没有文化差异
	5.从文化的角度来看，我和 AI 客服沟通很顺畅
	6.我与 AI 客服交流时不存在国家或民族的文化隔阂
G1. Always available	1. The AI chatbot is available in 7 days × 24 hours and at any device
	2. I can get the AI chatbot at anytime and anywhere
	3. The AI chatbot is always online
	4. The AI chatbot will never get off work
	1.AI 客服 7 天×24 小时提供服务
	2.我可以随时获得 AI 客服的服务
	3.AI 客服永远在线
	4.AI 客服永远不会下班
G2. Responsiveness	1. The AI chatbot responds quickly
	2. The AI chatbot is always ready to serve me
	3. The AI chatbot can always respond to me in time
	4. I feel that the AI chatbot is willing to serve me
	5. Every time I look for the AI chatbot, I always get a timely response
	1.AI 客服的回复速度很快
	2.AI 客服总是准备好了为我提供服务
	3.AI 客服总是能及时的响应我
	4.我感觉到 AI 客服乐于为我服务
	5.每次我找 AI 客服，总能及时得到回应
G3. Process simplify	1. The AI chatbot simplifies my service process
	2. The automatic service provided by the AI chatbot makes the online shopping easier
	3. The automatic service provided by the AI chatbot makes my problem solved more efficiently.
	1.AI 客服能够帮助我更快地获得想要的服务
	2.使用 AI 客服节约了我的时间
	3.AI 客服使我的问题得到更有效率的处理

Table 5 The results of structural model

Hypothesis	Path coefficient	p-value	Supported or not
H1a	0. 15***	<0. 001	Supported
H2a	0. 23***	<0. 001	Supported
H3a	0. 18***	<0. 001	Supported
H4a	0. 16***	<0. 001	Supported
H5a	0. 15***	<0. 001	Supported
H6a	0. 17***	<0. 001	Supported
H7a	0. 16***	<0. 001	Supported
H1b	0. 12***	<0. 001	Supported
H2b	0. 17***	<0. 001	Supported
H3b	0. 13***	<0. 001	Supported
H4b	0. 11**	=0. 009	Supported
H5b	0. 19***	<0. 001	Supported
H6b	0. 08*	=0. 033	Supported
H7b	0. 08*	=0. 034	Supported
H8	0. 22***	<0. 001	Supported
H9	0. 23***	<0. 001	Supported
H10	0. 29***	<0. 001	Supported

Note: *p<0.05；**p<0.01；***p<0.001.

Figure 1 The method roadmap

Phase 1: AICSQ- dimension development

- **Sample 1:** Organization sources, 55 organizations
- **Sample 2:** Customer sources, 47 customers
- **Dimension development:** Intergrating responses from different sources, iterating between practice and theory, using coding method of grounded theory

Phase 2: AICSQ-scale development

- **Sample 3:** 442 responses
- **Sample 4:** 471 responses
- **Scale development:** Define all the constructs, generate a sample of items, content validity test, develop a measurement model, collect first-round survey data, scale purification, collect second-round survey data, scale purification, cross-validation

Phase 3: Testing for nomological and predictive validity

- **Sample 5 :** 597 responses from platform users
- **Adopted constructs:** Satisfaction with AI chatbot, perceived value of AI chatbot, and intention of continous use
- **Estimation:** SEM with PLS estimates
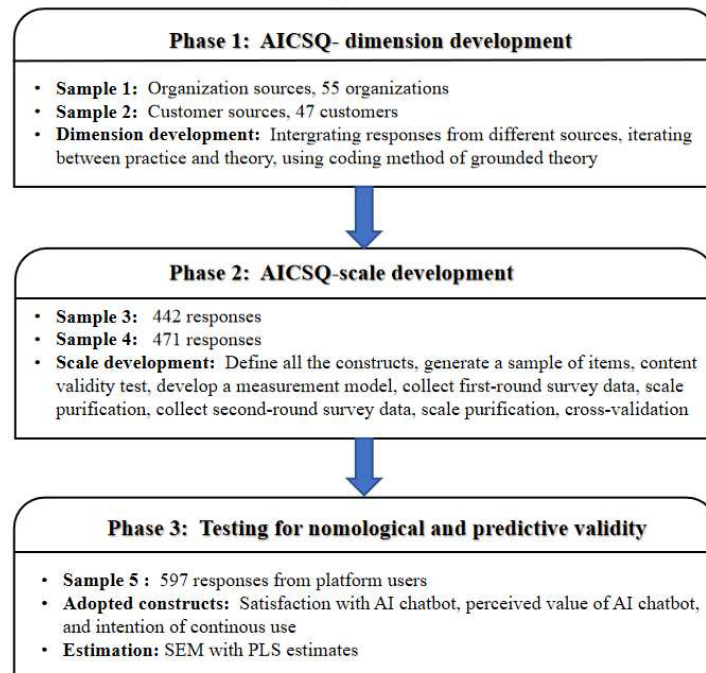
Figure 2   The multi-level dimension of AI chatbot quality

Semantic understanding	Close Human-AI Collaboration	Human-like	Continuous Improvement	Personalization	Cultural Adaption	Efficiency
• Understanding query • Understanding emotional expression	• Accessability • Easy to Transfer • Memorability	• Human-like Social Cue • Human-like Personality • Human-like Empathy	• Self-learning • System Update	• Identify Customers • Personalized Response • Personalized Recommendation	• Non-language Barriers • Culture Understanding	• Always Available • Responsiveness • Process simplify

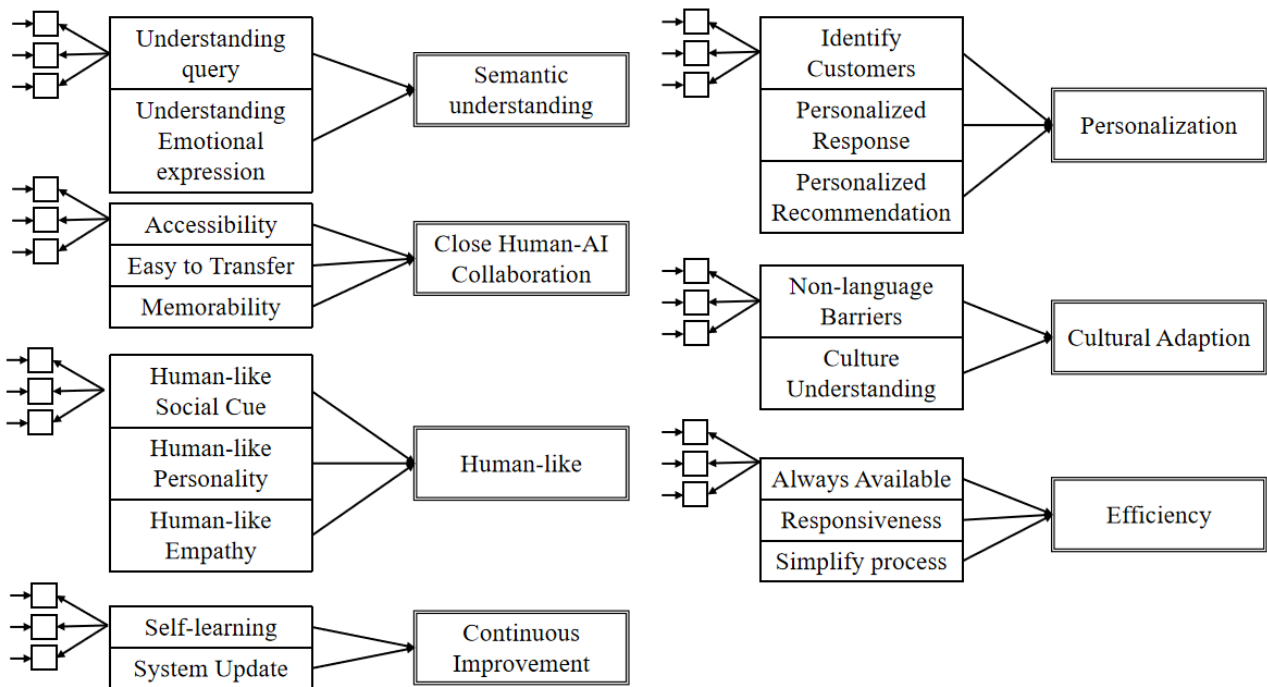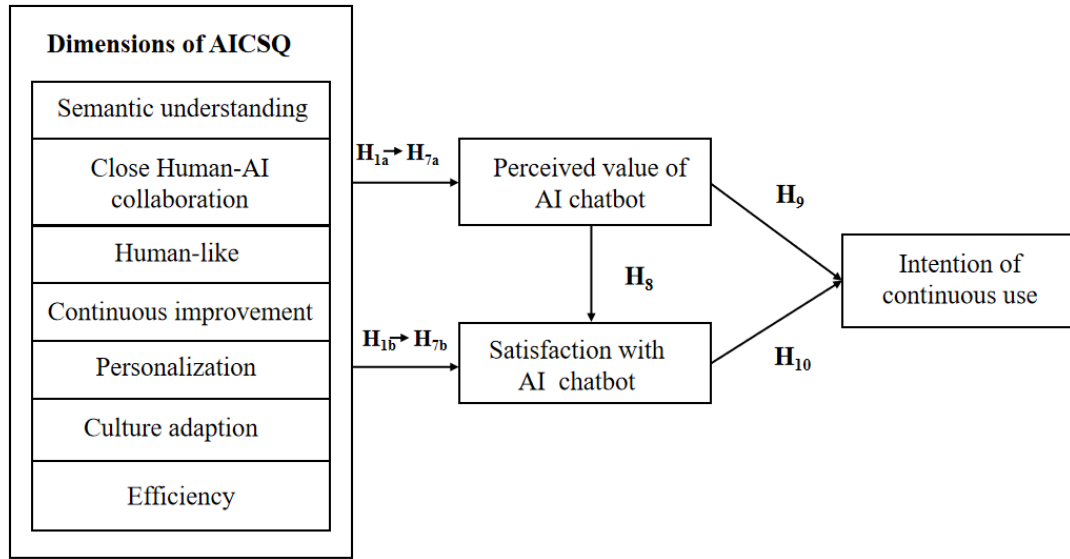## Figure 3 The measurement model of AI chatbot service quality

Figure 4   The research model for nomological test

Appendix A

The outline of the semi-structured questionnaire

1. Please describe AI Chatbot Service in your company.

2. What are the difference between an AI chatbot service and manual customer service?

3. Do you think your AI chatbot service is smart enough? If yes, why is it smart in your opinion? If not, why is not it smart?

4. How does your company coordinate the AI chatbot service and the manual customer service?

5. Can your customers turn to manual customer service by themselves? Under what condition will customers turn to manual customer service?

6. What are the elements and factors in the service quality of AI chatbot service?

7. What are the elements and factors your company would consider when designing AI chatbot?

8. Compared with other companies' AI chatbot service, what are the advantages of your company's?

9. Have your company received complaints from customers about AI chatbot service? What are these complaints?

10. Are customers satisfied with the AI chatbot service? What are the aspects why customers are dissatisfied or satisfied?

11. Are you familiar with other companies' AI chatbot service? Can you evaluate their merits and drawbacks?

12. Do you have other comments in this topic?

Reference

Anderson, J. C., & Gerbing, D. W. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology*, *76*(5), 732–740.

Arora, P., & Narula, S. (2018). Linkages between service quality, customer satisfaction and customer loyalty : A literature review. *Journal of Marketing Management*, *17*(4), 30–53.

Ashfaq, M., Yun, J., Yu, S., & Loureiro, S. M. C. (2020). I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. *Telematics and Informatics*, *54*(July), 101473.

Ba, S., Stallaert, J., & Zhang, Z. (2010). Balancing IT with the human touch: Optimal investment in IT-based customer service. *Information Systems Research*, *21*(3), 423–442.

Bagozzi, R., & Yi, Y. (1998). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94.

Barnes, S.J., & Vidgen, R. T. (2001). An evaluation of cyber-bookshops: The WebQual method. *International Journal of Electronic Commerce*, *6*, 6–25.

Barnes, Stuart J., & Vidgen, R. T. (2002). An integrative approach to the assessment of e-commerce quality. *Journal of Electronic Commerce Research*, *3*(3), 114–127.

Barrett, M., Davidson, E., Prabhu, J., & Vargo, S. L. (2015). Service innovation in the digital age: Key contributions and future directions. *MIS Quarterly*, *39*(1), 135–154.

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation confirmation model. *MIS Quartely*, *25*(3), 351–370.

Cai, S., & Jun, M. (2003). Internet users' perceptions of online service quality: A comparison of online buyers and information searchers. *Managing Service Quality*, *13*(6), 504–519.

Cenfetelli, R. T., Benbasat, I., & Al-Natour, S. (2008). Addressing the what and how of online services: Positioning supporting-services functionality and service quality for business-to-consumer success. *Information Systems Research*, *19*(2), 161–181.

China Economic and Information Research Center. (2019). *2019 China Artificial Intelligence Industry Research Report*.

Chung, M., Ko, E., Joung, H., & Kim, S. J. (2018). Chatbot e-service and customer satisfaction regarding luxury brands. *Journal of Business Research*, *117*, 587–595.

Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: an experimental study of human–chatbot interaction. *Future Generation Computer Systems*, *92*, 539–548.

Collier, J. E., & Kimes, S. E. (2012). Only If It Is Convenient : Understanding How Convenience Influences Self-Service Technology Evaluation. *Journal of Service Research*, *16*(1), 39–51.

Corbin J., & Strauss A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Qualitative Sociology*, *13*(1), 3–12.

Cox, J., & Dale, B. G. (2001). Service quality and e-commerce: an exploratory analysis. *Managing Service Quality*, *11*(2), 121–131.

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.

Cronin, J. ., Brady, M. ., & Hult, G. T. . (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing*, *76*, 193–218.

D'Mello, S. K., Graesser, A., & King, B. (2010). Toward spoken human-computer tutorial dialogues. *Human-Computer Interaction*, *25*(4), 289–323.

Dabholkar, P. A., Shepherd, C. D., & Thorpe, D. I. (2000). A comprehensive framework for service quality: An investi gation of critical conceptual and measurement issues through a longitudinal study. *Journal of Retailing*, *76*(2), 139–173.

De Keyser, A., Köcher, S., Alkire (née Nasr), L., Verbeeck, C., & Kandampully, J. (2019). Frontline service technology infusion: Conceptual archetypes and future research directions. *Journal of Service Management*, *30*(1), 156–183.

Deloitte. (2018a). *Digital labor intelligent chatbots*. https://www2.deloitte.com/us/en/pages/public-sector/articles/a-roadmapfor-building-digital-labor.html%0D

Deloitte. (2018b). *What is the business value of a chatbot?* https://www.deloitteforward.nl/en/digital/what-is-the-business-value-of-a-chatbot/

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

DeLone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, *19*(4), 9–30.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282.

Donmez, B., Pina, P. E., & Cummings, M. L. (2009). Evaluation criteria for human-automation performance metrics. In *Performance Evaluation and Benchmarking of Intelligent Systems* (pp. 21–40).

Forbes. (2017). *Ten customer service and customer experience trends for 2017*. https://www.forbes.com/sites/shephyken/2017/01/07/10-customer-service-and-customer-experience-cx-trends-for-2017/#145a272c75e5

Gartner. (2019). *How to manage customer service technology innovation*. https://www.gartner.com/smarterwithgartner/27297-2/

Gefen D, & Straub D. (2005). A practical guide to factorial validity using PLS-Graph: tutorial and annotated example. *Communications of the Association for Information Systems*, *16*(1), 91–109.

Golder, P. N., Mitra, D., & Moorman, C. (2012). What is quality? An integrative framework of processes and states. *Journal of Marketing*, *76*(4), 1–23. https://doi.org/10.1509/jm.09.0416

Gotlieb, J. ., Grewal, D., & Brown, S. . (1994). Consumer satisfaction and perceived quality: complementary or divergent constructs? *Journal of Applied Psychology*, *79*, 875–885.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619.

Harris, L. C., & Goode, M. M. H. (2004). The four levels of loyalty and the pivotal role of trust: A study of online service dynamics. *Journal of Retailing*, *80*(2), 139–158.

Hill, J., Ford, W. R., & Farreras, I. G. (2015). Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations.

Computers in Human Behavior, *Computers*(49), 245–250.

ISO 8373:2012—Robots and Robotic Devices (2012), (accessed October 22, 2021), available at https://www.iso.org/obp/ui/#iso:std:iso:8373:ed-2:v1:en

Jiang, J. J., Klein, G., & Carr, C. L. (2002). Measuring information system service quality: SERVQUAL from the other side. *MIS Quarterly*, *26*(2), 145–166.

Jörling, M., Böhm, R., & Paluch, S. (2019). Service robots: Drivers of perceived responsibility for service outcomes. *Journal of Service Research*, *22*(4), 404–420.

Kim, S. S. & Son, J. Y. (2009). Out of Dedication or Constraint? A Dual Model of Post-Adoption Phenomena and Its Empirical Test in the Context of Online Services. *MIS Quarterly*, *33*(1), 49–70.

Korfiatis, N., Stamolampros, P., Kourouthanassis, P., & Sagiadinos, V. (2019). Measuring service quality from unstructured data: A topic modeling application on airline passengers' online reviews. *Expert Systems with Applications*, *116*, 472–486.

Lacka, E., & Chong, A. (2016). Usability perspective on social media sites' adoption in the B2B context. *Industrial Marketing Management*, *54*, 80–91.

Lee, D., Oh, K. J., & Choi, H. J. (2017). The chatbot feels you - A counseling service using emotional response generation. *2017 IEEE International Conference on Big Data and Smart Computing, BigComp 2017*, 437–440.

Lee, H., & Lyu, J. (2016). Computers in Human Behavior Personal values as determinants of intentions to use self-service technology in retailing. *Computers in Human Behavior*, *60*, 322–332.

Li, X., & Sung, Y. (2021). Anthropomorphism brings us closer : The mediating role of psychological distance in User – AI assistant interactions. *Computers in Human Behavior*, *118*, 109.

Liang, H., Saraf, N., Hu, Q., & Xue, Y. (2007). Assimiliation of enterprise systems: The effect of institutional pressures and the mediating role of top management. *MIS Quartely*, *31*(1), 59–87.

Liu, C., & Arnett, K. . (2000). Exploring the factors associated with Web site success in the context of electronic commerce. *Information & Management*, *28*, 23–33.

Loureiro, S. M. C., Guerreiro, J., & Tussyadiah, I. (2021). Artificial intelligence in business: State of the art and future research agenda. *Journal of Business Research*, *129*(November 2020), 911–926.

Loureiro, S. M. C., Japutra, A., Molinillo, S., & Bilro, R. G. (2021). Stand by me: analyzing the tourist–intelligent voice assistant relationship quality. *International Journal of Contemporary Hospitality Management*.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *11*, 1–11.

Lusch, R. F., & Vargo, S. L. (2014). Service-dominant logic: Premises, perspectives, possibilities. In *Service-Dominant Logic: Premises, Perspectives, Possibilities*. Cambridge University Press.

Mackenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*(2), 293–334.

Madu, C. N., & Madu, A. A. (2002). Dimensions of e-quality. *International Journal of Quality & Reliability Management*, *19*(3), 246–258.

Maklan, S., Antonetti, P., & Whitty, S. (2017). A better way to manage customer experience: lessons from the royal bank of Scotland. *California Management Review*, *59*(2), 92–115.

Mani, Z., & Chouk, I. (2018). Consumer resistance to innovation in services: Challenges and barriers in the internet of things Era. *Journal of Product Innovation Management*, *35*(5), 780–807.

Marinova, D., de Ruyter, K., Huang, M. H., Meuter, M. L., & Challagalla, G. (2017). Getting smart: Learning from technology-empowered frontline interactions. *Journal of Service Research*, *20*(1), 29–42.

Melissa Archpru Akaka, & Stephen L. Vargo. (2015). Extending the context of service : From encounters to ecosystems. *Journal of Services Marketing*, *29*(6/7), 453–462. https://doi.org/10.1108/JSM-03-2015-0126

Mittal, V., Kumar, P., & Tsiros, M. (1999). Attribute-level performance, satisfaction, and behavioral intentions over time: A consumption-system approach. *Journal of Marketing*, *63*(2), 88–101.

Molla, A., & Licker, P. . (2001). E-commerce systems success: an attempt to extend and respecify the DeLone and McLean model of IS success. *Journal of Electronic Commerce Research*, 2, 1–11.

Mou, Y., & Xu, K. (2017). The media inequality: comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior*, *72*, 432–440.

Mourey, J. A., Olson, J. G., & Yoon, C. (2017). Products as pals: Engaging with anthropomorphic products mitigates the effects of social exclusion. *Journal of Consumer Research*, *44*(2), 414–431.

Murtarelli, G., Gregory, A., & Romenti, S. (2021). A conversation-based perspective for shaping ethical human – machine interactions : The particular challenge of chatbots. *Journal of Business Research*, *129*(March 2019), 927–935.

Novak, T. P., & Hoffman, D. L. (2019). Relationship journeys in the internet of things: a new framework for understanding interactions between consumers and smart objects. *Journal of the Academy of Marketing Science*, *47*(2), 216–237.

Nunnally, J. C. (1978). Psychometric theory. In *McGraw-Hill* (2nd ed.). McGraw-Hill.

O'Neill, M., Wright, C., & Fitz, F. (2001). Quality evaluation in online service environments: An application of the importance-performance measurement technique. *Managing Service Quality*, *11*(6), 402–417.

Obert, C., Probst, T. M., Martocchio, J. J., Drasgow, F., & Lawler, J. J. (2000). Empowerment and continuous improvement in the United States, Mexico, Poland, and India: Predicting fit on the basis of the dimensions of power distance and individualism. *Journal of Applied Psychology*, *85*(5), 643–658.

Oliver, R. L. (1981). Measurement and evaluation of satisfaction processes in retail settings. *Journal of Retailing*, *57*(3), 25–48.

Ostrom, A. L., Fotheringham, D., & Bitner, M. J. (2019). Customer acceptance of AI in service encounters: Understanding antecedents and consequences. In *Handbook of Service Science: Vol. II* (pp. 77–103). Springer Nature.

Overgoor, G., Chica, M., Rand, W., & Weishampel, A. (2019). Letting the computers take over: Using Ai to solve marketing problems. *California Management Review*, *61*(4), 156–185.

Palmer, J. W. (2002). Web Site Usability, Design and Performance Metrics. *Information Systems*

Research, *13*(2), 151–167.

Parasuraman, A., & Grewal, D. (2000). The impact of technology on the quality-value-loyalty chain: a research agenda. *Journal of the Academy of Marketing Science*, *28*, 168–174.

Parasuraman, A., Zeithaml, V. A., & Malhotra, A. (2005). E-S-QUAL a multiple-item scale for assessing electronic service quality. *Journal of Service Research*, *7*(3), 213–233.

Parasuraman, A., Zeithaml, V., & Berry, L. (1988). SERVQUAL:A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, *61*(1), 12–40.

Piercy, N. (2014). Online service quality: Content and process of analysis. *Journal of Marketing Management*, *30*(7–8), 747–785.

Qiu, L., & Benbasat, I. (2009). Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems. *Journal of Management Information Systems*, *25*(4), 145–182.

Rijsdijk, S. A., Hultink, E. J., & Diamantopoulos, A. (2007). Product intelligence: its conceptualization, measurement and impact on consumer satisfaction. *Journal of the Academy of Marketing Science*, *35*(3), 340–356.

Schroll, R., Schnurr, B., & Grewal, D. (2018). Humanizing Products with Handwritten Typefaces. *Journal of Consumer Research*, *45*, 648–673.

Sheehan, B., Jin, H. S., & Gottlieb, U. (2020). *Customer service chatbots : Anthropomorphism and adoption*. *115*(April), 14–24.

Shubin Lance Yu and Ji Jill Xiong (2019), How Chatbot Service Agents Can Alleviate the Negative Effect of Unresolved Requests on Consumers' Trust Toward Companies. In Rajesh Bagchi, Lauren Block, and Leonard Lee, Duluth (Eds.), *NA - Advances in Consumer Research Volume 47*(PP: 926-927), MN : Association for Consumer Research.

Tan, C.-W., Benbasat, I., & Cenfetelli, R. T. (2013). IT-mediated customer service content and delivery in electronic governments: an empirical service quality. *MIS Quarterly*, *37*(1), 77–109.

Tarn, K. Y., & Ho, S. Y. (2005). Web Personalization as a Persuasion Strategy: An Elaboration Likelihood Model Perspective. *Information Systems Research*, *16*(3), 271–291.

Thomas, D. M., & Bostrom, R. P. (2010). Vital signs for virtual teams: An empirically developed trigger model for technology adaptation interventions. *MIS Quarterly*, *34*(1), 115–142.

Thomaz, F., Salge, C., Karahanna, E., & Hulland, J. (2020). Learning from the dark web: leveraging conversational agents in the era of hyper-privacy to enhance marketing. *Journal of the Academy of Marketing Science*, *48*(1), 43–63.

Tian, Y., Chen, X., Xiong, H., Li, H., Dai, L., Chen, J., Xing, J., Chen, J., Wu, X., Hu, W., Hu, Y., Huang, T., & Gao, W. (2017). Towards human-like and transhuman perception in AI 2.0: A review. *Frontiers of Information Technology & Electronic Engineering*, *18*(1), 58–67.

Wang, Y. S. (2008). Assessing e-commerce systems success: A respecification and validation of the DeLone and McLean model of IS success. *Information Systems Journal*, *18*(5), 529–557.

Watkins, L., & J. Gnoth. (2011). The value orientation approach to understanding culture. *Annals of Tourism Research*, *38*(4), 1274–1299.

Watson, R. T., Pitt, L. F., & Kavan, C. B. (2006). Measuring Information Systems Service Quality: Lessons from Two Longitudinal Case Studies. *MIS Quarterly*, *22*(1), 61.

Weber, R. P. (1990). *Basic Content Analysis*. Sage.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931.

Wixom, B. H., & Todd, P. A. (2005). A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*, *16*, 85–102.

Wolfinbarger, M., & Gilly, M. C. (2003). eTailQ: Dimensionalizing, measuring and predicting etail quality. *Journal of Retailing*, *79*(3), 183–198.

Xu, C., Peak, D., & Prybutok, V. (2015). A customer value, satisfaction, and loyalty perspective of mobile application recommendations. *Decision Support Systems*, *79*, 171–183.

Yang, Z., & Peterson, R. T. (2004). Customer perceived value, satisfaction, and loyalty:the role of switching costs. *Psychology and Marketing*, *21*(10), 799–822.

Yoo, B., & Donthu, N. (2001). Developing a scale to measure the perceived quality of an internet shopping site (SITEQUAL). *Quarterly Journal of Electronic Commerce*, *2*(1), 31–47.

You, S., & Robert, L. P. (2018). Emotional Attachment , Performance , and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots. *Journal of the Association for Information Systems*, *19*(5), 377–407.

Zeithaml V A, Parasuraman A, & Malhotra A. (2005). A conceptual framework for understanding e-service quality: Implications for future research and managerial practice. *Journal of Service Research*, *7*, 1–21.