Future Technology
Open Access Journal
ISSN 2832-0379

November 2026| Volume 05 | Issue 04 | Pages 34-48

Journal homepage: https://fupubco.com/futech

https://doi.org/10.55670/fpll.futech.5.4.4

Future Publishing LLC

Article

Evaluating human-like chatbot interaction in customer service: integrating anthropomorphism and conversational capability

Warjiyono Warjiyono[1]*, Hadiyanto Hadiyanto[2], Dinar Mutiara Kusumo Nugraheni[3]

[1]Doctoral Program of Information System, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia
[2]Department of Chemical Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia
[3]Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

ARTICLE INFO

Article history:
Received 28 January 2026
Received in revised form
14 June 2026
Accepted 18 July 2026

Keywords:
AI chatbot, Customer service, Anthropomorphism, Human-Like Chatbot Interaction, Conversational capability

*Corresponding author
Email address:
warjiyono@students.undip.ac.id

DOI: 10.55670/fpll.futech.5.4.4

ABSTRACT

The development of artificial intelligence (AI)-based technology has driven the development of rule-based chatbots into AI-based chatbots. This development is used by various companies to use AI chatbots as the frontline of customer service. Meeting user expectations for AI chatbots requires an evaluation model to measure chatbot quality. Based on the literature, chatbot evaluation models still focus solely on technical quality, even as AI chatbots are evolving towards human-like services. This research will explore the integration of the technical dimensions of information system success with a new dimension, namely Human-Like Chatbot Interaction (HLCI) as a higher-order construct of anthropomorphism and conversational capability. This quantitative study uses an explanatory research approach. Data were collected through a cross-sectional survey of users of the Veronika chatbot, owned by Telkomsel, an Indonesian telecommunications company. A total of 213 respondents had data ready for analysis using the Partial Least Squares Structural Equation Modeling (PLS-SEM) method in SmartPLS 4, namely the measurement model and structural model tests. The analysis results show that service quality is not significantly related to user satisfaction but has a direct effect on trust. Meanwhile, information quality, system quality, and human-like chatbot interaction are positively related to trust and user satisfaction. Trust and user satisfaction are positively related to intention to continue using. The results of this study can contribute both theoretically and practically to modern AI chatbot quality evaluation models.

1. Introduction

Artificial intelligence (AI) technology has driven a shift in customer service from human-based to chatbot-based services. Advances in Natural Language Processing (NLP), Machine Learning (ML), and large language models (LLM) have transformed chatbots from rule-based systems into chatbots capable of delivering conversations that mimic human interaction. AI chatbots offer numerous benefits, including anytime, anywhere accessibility, reduced operational costs, and the ability to provide accurate, accountable information from document databases [1]. These advantages make many companies in various sectors such as banking, e-commerce, healthcare, and government [2,3]. Companies have shifted to using AI-based chatbots as communication channels between companies and customers. The presence of AI chatbots as a medium for interaction between companies and customers has shifted user expectations from simply receiving information to a conversational experience resembling human interaction. Despite the advancement of AI chatbots in customer service, the conceptual framework for evaluating AI chatbots has yet to adapt to the advancements in modern AI technology. Current AI chatbot evaluation models still adopt traditional information systems success theories that focus solely on technical factors such as system, service, and information quality, without incorporating social interaction, which is widely recognized as a key factor in the success of user experiences with AI chatbots [4]. Digital service quality should not only focus on technical quality but also expand to include chatbots that can provide human-like interactions. AI chatbot evaluation models that emphasize technical factors alone cannot assess the quality of modern AI chatbots that prioritize social interaction, natural communication, and empathy. Although AI chatbots are increasingly used for customer service across companies, many studies indicate that many AI chatbots still provide stiff, unnatural communication, lack empathy, and are unable to build interpersonal relationships with users. This condition leads to low user trust, satisfaction, and intention to use the chatbot again. Many previous studies have shown that trust is a key factor in AI chatbot services; without it, users are reluctant to use AI chatbots again [5]. Today, the success of a modern chatbot is no longer determined solely by technical capabilities but also by the quality of the user's interaction experience.

Recent research has begun to explore the importance of non-technical factors such as anthropomorphism and human-like conversational abilities in determining the success of AI chatbots. This is in line with previous research [6] indicating that chatbots with anthropomorphic elements can positively influence trust formation, user satisfaction, and intention to reuse AI chatbots [7]. Meanwhile, chatbots with human-like conversations can foster user engagement, and the sense of social presence during interaction will encourage the intention to reuse the chatbot [8]. Various studies have shown that these two constructs influence trust formation, user satisfaction, and the intention to reuse chatbots. However, most studies still consider these two constructs as independent entities. Separating anthropomorphism and conversational capability leads to a less comprehensive understanding of chatbot interactions. It's like humans can only be friendly and polite, but cannot communicate in a well-structured, understandable manner. Therefore, these two dimensions conceptually complement each other in shaping user perceptions of human-like interactions. To date, there remains a theoretical gap: no unified construct represents the characteristics and interactions of modern AI chatbots. Both variables, anthropomorphism and conversational capability, can influence user trust and satisfaction [9].

To address this gap, this study introduces a new dimension, namely Human-Like Chatbot Interaction, which integrates the Anthropomorphism and Conversational Capability dimensions into a higher-order construct. The development of this new dimension, namely Human-Like Chatbot Interaction, is supported by the Computers Are Social Actors (CASA) theory [10], which introduces the understanding that interactions with computers that provide social styles, such as presenting human-like characters with language styles similar to humans, will be considered by computers as social actors. The next theory is Social Presence Theory [11], which explains that during an interaction, the perceived quality of communication is influenced by social presence, and it is suitable for chatbots with AI technology, not desktop computers [12]. Social Presence Theory also reveals that anthropomorphism features in chatbots can influence user responses, as evidenced by humanoid social robots [8]. Supported by ecommerce research showing that chatbots have human-like characteristics and conversational qualities that can increase trust and encourage users to use the chatbot again [13]. Research in the fintech sector also reveals that chatbots exhibit human-like characteristics and conversational behaviors that influence users' purchase intentions [14]. Therefore, separating the dimensions of Anthropomorphism and Conversational Capability will reduce the holistic nature of AI chatbot interactions. The development of this model provides a theoretical contribution, namely expanding the traditional information system success model by introducing a new construct, namely Human-Like Chatbot Interaction as a higher-order construct of Anthropomorphism and Conversational Capability. The practical contribution is to provide a more comprehensive evaluation framework for organizations to design AI chatbots that increase trust, user satisfaction, and continued use. The methodological contribution is to validate the higher-order construct in the AI chatbot success model through empirical testing.

2. Literature Review

2.1 AI chatbot in customer services

In the past, chatbots were first used for rule-based customer service because AI technology didn't exist at the time. However, over time, with the advent of AI, customer service has shifted toward AI-based chatbots. These AI-based chatbots are capable of providing real-time customer service, accessible anywhere, anytime, and accurate, up-to-date information, thanks to their ability to read document databases. AI chatbots also benefit companies by increasing efficiency, reducing operational costs, improving customer service effectiveness, and adding value. With the advantages of AI-based chatbots, companies now widely use them as automated services for customers, such as general information, product information, questions, and other technical services. With the increasing use of AI chatbots in companies, many chatbots are still created without considering quality and user satisfaction. In their interactions, chatbots appear stiff, unnatural, and unfriendly, failing to meet user expectations and resulting in low user trust and satisfaction [15]. This shows that the success of a chatbot depends not only on technical capabilities but also on the quality of the conversation, making the AI chatbot more comfortable to use. The theoretical concept of Human–Computer Interaction (HCI) is that user experience is paramount in the interaction between humans and computers, in addition to the ease and usability of the computer system. Modern HCI concepts in AI chatbots emphasize the quality of AI chatbots, which provide friendly and responsive conversations similar to interactions with human service [16].

2.2 Relevance of CASA theory to chatbots

Computers Are Social Actors (CASA) theory explains that computer responses that include social cues will describe the service being performed by humans [17]. However, some theories, such as Media Equation Theory [10], state that individuals treat media the same way they treat humans. Social Presence Theory [11] explains that during an interaction, the perceived quality of communication is influenced by social presence. Parasocial Interaction Theory [18] explains that, even though the interaction is one-way, users feel emotional closeness and a social connection with the media. CASA Theory is considered particularly appropriate for AI chatbots. As stated by [19], users of digital agents will adhere to social rules and expectations of digital agents that possess human-like characteristics. Parasocial Interaction Theory focuses on a one-way relationship between users and the medium, while AI chatbots are interactive, two-way communication [18]. Therefore, CASA Theory provides a more appropriate theoretical foundation for understanding anthropomorphic perceptions of AI chatbots [20]. Previous research has shown that chatbots with human-like communication styles increase perceptions of social presence, as users feel emotionally and socially comfortable with the chatbot [21], which in turn can increase trust and satisfaction [22]. In this study, Social Presence Theory serves as the theoretical basis for the Conversational Capability dimension.

Anthropomorphism in AI chatbots: AI chatbots with anthropomorphic features will appear more human, exhibiting expressions, empathy, and emotional intelligence similar to those of characters in human-provided services [23]. This condition will influence user satisfaction [24], trust [25], and the user's intention to use the chatbot again in the future when they need services.

Conversational capability in AI chatbots: The concept of conversational capability in chatbots describes chatbots as a communication medium with customers that can support natural dialogue and structured conversations and can understand human language even when questions contain incorrect words or non-standard sentence structure [26,27]. Previous research has revealed that the future of AI chatbot design must prioritize conversation quality and the ability to handle repeated questions with consistent answers [23].

Dimensions in AI chatbots: The dimensions that will be used in this study are sourced from several previous theories that have been widely used and trusted by previous researchers in measuring the quality of service, namely the SERVQUAL model [28] and the theory of measuring the quality of e-commerce, namely the Delone and McLean model [29] and the theory of measuring the quality of health information systems, namely the Human Organization Technology Fit (HOT-Fit) model [30]. These three models are the main foundation for evaluating the quality of AI chatbots in customer service.

2.3 Development of human-like chatbot interaction

The development of Artificial Intelligence (AI)-based chatbots enables them to understand conversational context, generate dynamic responses, and exhibit human-like communication characteristics (Table 1). This development demonstrates that the success of AI chatbots is no longer determined solely by the system's technical quality, but also by the quality of the interactions that create a natural communication experience.

Based on Computers Are Social Actors (CASA) Theory [17] and Social Presence Theory [11], CASA Theory posits that users treat technology that displays social cues similarly to how they treat human social interactions. Meanwhile, Social Presence Theory posits that the greater a medium's ability to provide social presence, the more users perceive the interaction as human-like. Both theories provide a basis for the view that the experience of human interaction with AI chatbots is shaped not only by the system's technical capabilities but also by the chatbot's ability to exhibit human-like social and communicative characteristics. This research develops the construct of Human-Like Chatbot Interaction (HLCI) as a Higher-Order Construct formed by Anthropomorphism and Conversational Capability. Anthropomorphism represents the user's perception that the chatbot possesses human-like characteristics, such as friendliness, politeness, and empathy. Conversational Capability represents the chatbot's ability to engage in natural, structured conversation. This integration is based on the belief that user perception of human-like interactions is a holistic experience and cannot be explained through a single dimension (Figure 1).

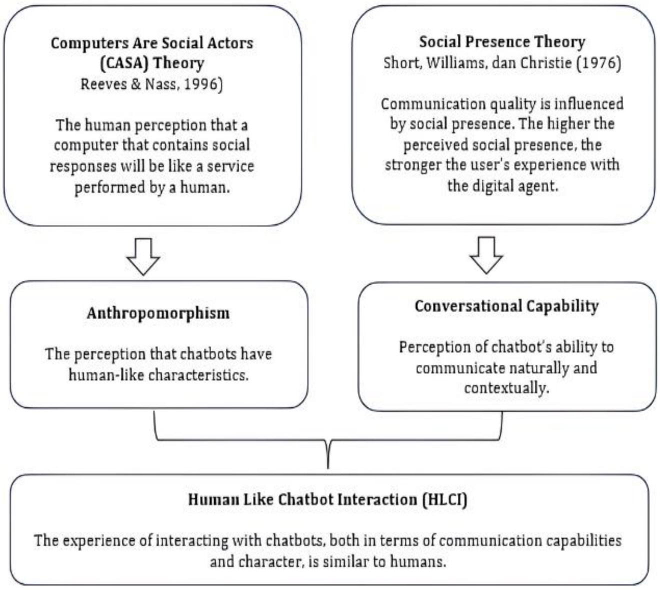

Figure 1. Theoretical justification of human-like chatbot interaction

Table 1. Dimensions in AI chatbots

Dimensions	Information	Ref.
System Quality	Measuring system performance, such as system performance, access time, speed, ease of use, and access at any time.	[28,29,31]
Information Quality	Measuring the quality of the information produced by the system. Is the information produced accurate, up to date, complete, and clear to understand?	[28,29,31]
Service Quality	Measuring the quality of service provided by the system. Is the service provided by the system responsive, able to provide support when problems arise, and comfortable to use?	[28,29,31]
Trust	Measuring user trust reactions to systems that have provided services.	[33,34,35]
User Satisfaction	Measuring user satisfaction after using the system. This satisfaction is based on how the system's expectations align with reality and the user's desires.	[30,32,33,36]
Intention of Continuous Use	Measuring users' desire and intention to reuse the system when needed	[31,32,34]

2.4 Building hypotheses

A summary of the proposed research hypothesis is given in Figure 2.

H1: Service quality has a positive effect on user satisfaction. The relationship between Service Quality and User Satisfaction is based on SERVQUAL Theory [28] and the DeLone and McLean IS Success Model [29]. Good service quality enables users to receive responsive, reliable, and tailored assistance, thereby increasing positive evaluations of the user experience. In the context of AI chatbots, high service quality enables users to obtain solutions more effectively and efficiently, resulting in higher satisfaction. Several previous studies have shown that service quality is a key determinant of chatbot user satisfaction [36].

H2: Information quality has a positive effect on user satisfaction. The relationship between Information Quality and User Satisfaction is based on the DeLone and McLean IS Success Model [29]. Accurate, relevant, complete, and easy-to-understand information enhances the user experience and results in more positive system evaluations. In Generative AI chatbots, good information quality enables users to obtain answers tailored to their needs, thereby increasing user satisfaction. This finding is supported by research [37].

H3: System Quality has a positive effect on User Satisfaction. The relationship between System Quality and User Satisfaction is based on the DeLone and McLean IS Success Model [29]. High system quality, such as ease of use, responsiveness, and reliability, enables users to experience a better experience, thereby increasing user satisfaction. This influence has been demonstrated in various information systems and AI chatbot contexts [38].

H4: Human-like chatbot interaction has a positive effect on user satisfaction.

The relationship between human-like chatbot interaction and user satisfaction is based on Computers Are Social Actors (CASA) theory [10] and social presence theory [11]. Users tend to respond positively to chatbots that exhibit human-like characteristics and communicate naturally. Human-like interactions enhance user comfort and experience, resulting in higher levels of satisfaction [39].

H5: Service quality has a positive influence on trust.

According to the SERVQUAL Theory [28] and the DeLone and McLean IS Success Model [29], good service quality can increase user confidence in the service provider's ability to meet their needs. In the context of AI chatbots, a responsive, reliable service that resolves user issues will increase perceptions of trust in the chatbot. Previous research has shown that service quality positively influences trust in digital services and AI chatbots [40].

H6: Information quality has a positive effect on trust.

According to the DeLone and McLean IS Success Model [29], good information quality will increase user confidence in the system. In the context of AI chatbots, accurate and relevant information will reduce uncertainty and increase user perceptions of trust in the chatbot. Research by [41] shows that information quality is a crucial factor in building trust.

H7: System quality has a positive effect on trust.

Good system quality will increase user confidence in the system's ability to consistently provide services. A system that is easy to use, stable, and error-free will reduce perceived risk, thereby increasing user trust. This relationship has been demonstrated across various digital platforms and AI-based services [42].

H8: Human-like Chatbot Interaction has a positive effect on Trust.

According to CASA Theory [10] and Social Presence Theory [11], the perceived social presence and human-like characteristics of a chatbot can enhance perceptions of competence and social closeness, thereby strengthening user trust. The greater a chatbot's ability to provide a human-like interaction experience, the higher the user's level of trust in the chatbot. This relationship is supported by research by [43].

H9: Trust has a positive effect on user satisfaction.

The relationship between trust and user satisfaction is based on trust theory [44]. Trust is a psychological mechanism that can reduce uncertainty and increase positive evaluations of technology use experiences. In the context of AI chatbots, users who trust chatbots tend to feel more comfortable and satisfied during interactions. Research by [45] shows that trust is an important predictor of user satisfaction.

H10: User Satisfaction has a positive effect on Intention of Continuous Use.

The relationship between User Satisfaction and Intention to Continue Using is based on Expectation-Confirmation Theory [46]. This theory explains that users who experience a system that meets their expectations will be satisfied and likely to continue using it. In the context of AI chatbots, user satisfaction with the quality of interactions and services provided will increase their intention to continue using the system. This finding is supported by research by [47].

H11: Trust has a positive effect on Intention of Continuous Use.

The relationship between trust and Intention of Continuous Use a service is based on trust theory and expectancy confirmation theory [46]. User trust in a chatbot will reduce perceived risk and increase their confidence in continuing to use the service. Previous research indicates that trust is a crucial factor in the continued use of digital services [48].

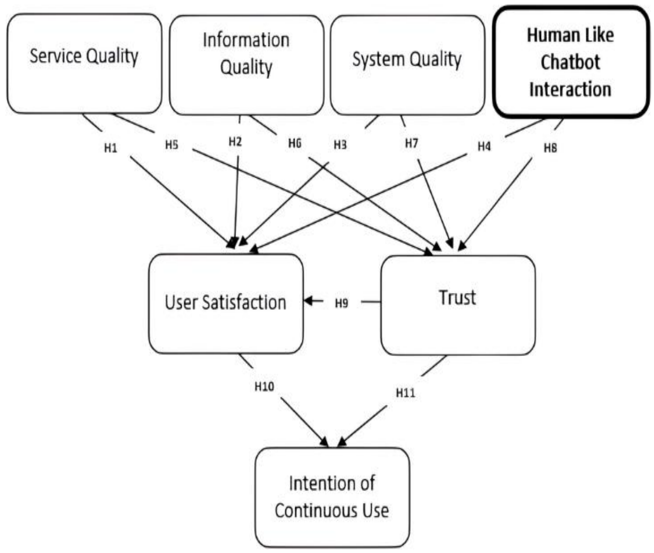

Figure 2. Proposed research hypothesis

3. Research methodology

3.1 Research design

This study employs a quantitative approach and an explanatory research design to examine the relationships among constructs in an AI chatbot evaluation model based on user experience. Explanatory research was chosen because it is widely used in information systems research to test models and is well-suited to this study [49]. The Indonesian telecommunications company Telkomsel's chatbot, Veronika, was chosen as the subject of this research because it is AI-based. Veronika chatbot users will be surveyed online in a cross-sectional study about their experiences.

3.2 Sampling purposive

A total of 300 respondents were selected based on the criterion of having used the Veronika chatbot at least once. To ensure this sampling purpose, the initial questionnaire asked whether or not they had used the Veronika chatbot. If not, the survey stopped; if yes, it continued. Purposeful sampling was suitable for this study because respondents were people who had used the chatbot being studied [50]. A total of 213 respondents stated that the data was good and avoided questionnaire content that showed a straightlining response pattern (i.e. the questionnaire content was the same across all items), which was considered an indication of careless responses [51].

A total of 213 respondents met the requirements for analysis using the Partial Least Squares Structural Equation Modeling (PLS-SEM) method. The sample size was determined based on the 10-fold rule [49], which is based on the number of paths entering the latent construct. The highest number of paths, namely User Satisfaction, received: H1, H2, H3, H4, and H9 = 5 paths, meaning 10 × 5 = 50. Therefore, the minimum number of respondents was 50. Furthermore, Ref [49] also recommends using G*Power software to generate more accurate minimum sample estimates. For this purpose, the sample size calculation was performed using G*Power software by selecting F tests, Linear multiple regression: Fixed model, R^2 deviation from zero, a priori power analysis, effect size f^2 of 0.15, α of 0.05, power of 0.80, and a maximum of 5 predictors. The calculation results require 92 respondents. Therefore, it can be concluded that 213 respondents have exceeded the minimum limit.

3.3 Respondent profile

The demographic profile of the 213 respondents, by gender, is predominantly male (74%), with 26% female. By age, the majority of respondents are aged 24–30 (29%), followed by those aged 31–39 (23%), those aged 18–23 (19%), those aged 40–49 (17%), those aged 17 and under (8%), those aged 50–59 (4%), and those aged 60 and over (1%). This indicates that the majority of users are in the productive and digitally active age group. Respondents, by education level, had the highest number of high school graduates (53%), followed by bachelor's degrees (29%), diplomas (9%), junior high school (5%), master's degrees (2%), and doctoral degrees (0%). This indicates that AI chatbot users are predominantly those with secondary and undergraduate degrees. Respondents, by occupation, were mostly private-sector employees (30.5%), freelancers or informal workers (23%), self-employed (22.5%), students (21.1%), civil servants (2.3%), and retirees (0.5%). This reflects the relevance of chatbot services to economically active users. Based on chatbot usage behavior, most respondents reported using the Veronika chatbot only once or twice in the previous six months (46.5%), while 39.4% used it three to five times, 10.8% used it six to ten times, and 3.3% used it more than ten times. This indicates that many users are relatively new.

Based on the duration of chatbot use, the majority of respondents had used the chatbot for less than one year (46.5%), one year (23.9%), two years (15%), or more than three years (14.6%). This indicates that most users are new, likely still exploring the features and benefits of chatbots. Based on the purpose of chatbot use, the most common uses are for information searches (70.9%), complaints (48.8%), product inquiries (33.3%), and personal use (2.3%). This indicates that chatbots are widely used to answer customer questions.

3.4 Common method bias test

All constructs in this study were measured using a single questionnaire completed by the same respondents, which can lead to Common Method Bias (CMB). Because all constructs were measured through a single survey, this study evaluated the potential for CMB using the Full Collinearity VIF approach [52]. To mitigate the CMB problem, statistical testing using a full collinearity approach is required, where all constructs are regressed on a common factor, and the variance inflation factor (VIF) value is calculated. According to [52], a VIF value ≤ 3.3 is considered free from Common Method Bias. In PLS-SEM, interpretation of results also takes into account the recommendation of [49], which states that a VIF value ≤ 5 does not indicate a serious collinearity problem. Therefore, this study uses two interpretations, namely VIF ≤3.3 as an indicator of the absence of Common Method Bias based on [52], while VIF ≤5 is used as an indicator that there are no serious multicollinearity problems according to the recommendations of [49].

Table 2 indicates that all constructs have VIF values below 5.0. Although several constructs exceed the conservative limit of 3.3 [52], namely Human-Like Chatbot Interaction, Information Quality, System Quality, and Trust, all values are still below the 5.0 limit recommended by [49] for evaluating collinearity in PLS-SEM models. Furthermore, a VIF threshold of <5.0 has been used in recent PLS-SEM studies, such as [53]. Therefore, the research model is considered free of serious collinearity problems and is suitable for proceeding to the measurement and structural model evaluation stages.

Table 2. Full collinearity VIF test

Construct	Full Collinearity VIF
Service Quality	3.248
System Quality	3.451
Information Quality	3.788
Human Like Chatbot Interaction	3.677
Trust	3,184
User Satisfaction	4.537
Intention of Continuous Use	2.887

3.5 Measurement development

The measurement instrument consists of seven dimensions and 31 indicators (Table 3). The dimensions to be measured are Service Quality (SQ1-SQ5), Information Quality (IQ1-IQ5), System Quality (SYQ1-SYQ6), Human-Like Chatbot Interaction (HLC1-HLC6), Trust (TR1-TR3), User Satisfaction (US1-US3), and Intention of Continuous Use (IC1-IC3). The instrument uses a five-point Likert scale (1 = strongly disagree to 5 = strongly agree). This scale was used because it was appropriate for this study, as it captured respondents' perceptions of the effectiveness of the Veronika chatbot system.

3.6 Data analysis technique

The 213 respondents' data, which have been declared valid, will be analyzed using the Partial Least Squares Structural Equation Modeling (PLS-SEM) method using SmartPLS 4 software. The data analysis will involve two stages (Figure 3): a measurement model test and a structural model test. The measurement model test stage assesses the validity and reliability of the constructs, while the structural model test examines the relationships between the constructs. The test stage is based on the Partial Least Squares Structural Equation Modeling (PLS-SEM) method [49].

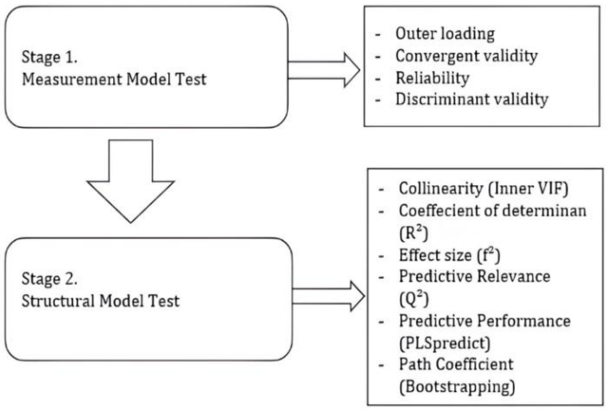

Figure 3. Testing stage

Table 3. Measurement instruments

Constructs	Items	Ref.
Service Quality (SQ)	SQ1: The chatbot interface makes me feel comfortable when interacting.	[28,29,31]
	SQ2: The chatbot provided a quick response to my questions.	[31,67,68]
	SQ3: I feel my personal data and information is safe when using chatbots.	[28,29,32,34]
	SQ4: When the chatbot cannot answer my question, the chatbot will provide information or an explanation.	[32,34]
	SQ5: The chatbot provided assistance when I was having trouble.	[47]
Information Quality (IQ)	IQ1: Chatbot provides accurate information.	[47]
	IQ2: The chatbot provides complete information to answer my needs.	[34]
	IQ3: The answer from the chatbot is relevant to my question	[34]
	IQ4: Chatbot provides up-to-date information	[33]
	IQ5: The information conveyed by the chatbot is easy for me to understand.	[33]
System Quality (SYQ)	SYQ1: Chatbot is easy for me to use	[29,30]
	SYQ2: The chatbot ran smoothly during my time using it.	[28,29,32]
	SYQ3: Chatbot can be accessed quickly	[34]
	SYQ4: I feel the chatbot is safe to use.	[31]
	SYQ5: I can access the chatbot whenever needed.	[47]
	SYQ6: Chatbots are always available when needed	[34,47]
Human Like Chatbot Interaction (HLCI)	HLCI1: The chatbot understands my feelings during the conversation.	[31,47]
	HLCI2: This chatbot provides greetings and greetings when interacting.	[31,47,55]
	HLCI3: Chatbots display responses that give the impression of having human-like expressions.	[31,47,55]
	HLCI4: Conversations with this chatbot feel as natural as talking to a human.	[32,33]
	HLCI5: Chatbot is able to understand the meaning of the questions I ask.	[32,33]
	HLCI6: The conversation flow with this chatbot runs smoothly.	[55]
User Satisfaction (US)	US1: The experience of interacting with the chatbot was in line with my expectations.	[29,34]
	US2: I found the experience of using the chatbot enjoyable.	[31–33]
	US3: Overall, I am satisfied using this chatbot.	[31,32,34,35]
Trust (TR)	TR1: I believe the chatbot provides consistent service every time I use it.	[33]
	TR2: I believe chatbots will help when needed	[32,33]
	TR3: I believe this chatbot is trustworthy.	[34]
Intention of Continuous Use (IC)	IC1: I intend to continue using this chatbot in the future.	[31,32,34]
	IC2: I am willing to recommend this chatbot to others.	[31]
	IC3: I intend to use this chatbot again when I need service.	[31,32,34]

4. Results

4.1 Measurement model test

The measurement model test will examine the validity and reliability of 31 indicators to determine whether each indicator measures its construct. There are four tests:

Outer loading test: An initial evaluation is conducted using an outer loading test to determine whether each indicator represents its construct. If the outer loading value is ≥0.70, the indicator is considered very good. If the outer loading value is 0.50–0.69, the indicator is retained. If the outer loading value is <0.50, the indicator must be removed [49]. The outer loading test of 31 indicators produced outer loading values between 0.712 -0.941 The outer loading test of 31 indicators produced outer loading values between 0.712 - 0.941. The lowest outer loading value was for indicator SYQ6, which stated that the chatbot is always available when needed, while the highest was for the chatbot providing a pleasant service experience. With outer loading values >0.70, all 31 indicators were considered very good at measuring their constructs and could be used for further testing of the structural model.

Convergent validity test: This test will assess 31 indicators, whether each indicator represents its construct or not. This provision uses the Average Variance Extracted (AVE) value, namely if the AVE value ≥ 0.50 then it is fulfilled, if the AVE value < 0.50, then the indicator is not fulfilled because it is less able to represent its construct [49]. The results of the convergent validity test in Table 4 show that all constructs have AVE values ≥ 0.50, ranging from 0.675 to 0.867. It is concluded that the 31 indicators represent the seven constructs and can be continued for the structural model test.

Reliability test: The reliability test was used to assess whether the 31 indicators are reliable for measuring the 7 constructs. The assessment uses Cronbach's Alpha and Composite Reliability. If the Cronbach's Alpha and Composite Reliability values are CR ≥ 0.70, it is declared reliable and acceptable; if CR ≥ 0.60 it is still acceptable, if CR < 0.60 it is declared unreliable [49]. The reliability test results in Table 5 show that the seven constructs have Cronbach's alpha values ranging from 0.889 to 0.923, and the composite reliability values range from 0.922 to 0.941. Since all constructs have Cronbach's alpha values ≥ 0.70, they are deemed reliable and acceptable for further structural model testing.

Table 4. Convergent validity test

Test	Average variance extracted (AVE)
Service Quality	0.702
Information Quality	0.726
System Quality	0.679
Human-Like Chatbot Interaction	0.675
Trust	0.818
User Satisfaction	0.867
Intention of Continuous Use	0.848

4.2 Discriminant validity test

The differences between constructs are then assessed using discriminant validity tests to determine whether each construct is distinct from the others. If the HTMT value is <0.85, the result is very good discriminant validity, if the HTMT value is <0.90, the result is acceptable discriminant validity, and if the HTMT value is >0.90, the result is not accepted because the constructs may be too similar. The results of the discriminant validity test in Table 6 show HTMT values between 0.693 and 0.884, indicating very good discriminant validity: each construct is distinct from the others and can be continued to structural testing.

Table 5. Reliability test

Test	Cronbach's alpha	Composite reliability (rho_c)
Service Quality	0.894	0.922
Information Quality	0.905	0.930
System Quality	0.904	0.927
Human Like Chatbot Interaction	0.904	0.926
Trust	0.889	0.931
User Satisfaction	0.923	0.941
Intention of Continuous Use	0.910	0.944

The results of the Higher-Order Construct evaluation in Table 7 show that Anthropomorphism, with a path coefficient (β) of 0.544 and a t-value of 46.816, and Conversational Capability (β = 0.536 and a t-value of 42.543, with an overall p-value ($p < 0.001$), significantly shape the construct of Human-Like Chatbot Interaction. All 95% bootstrap confidence intervals did not cross zero, indicating that both dimensions significantly contribute to the formation of the second-order construct. These evaluation results support the conceptualization that human-like chatbot interactions are determined not only by user perceptions of the chatbot's human characteristics but also by the chatbot's ability to build natural, responsive, and contextual conversations. All constructs meet the criteria for internal consistency reliability and convergent validity. Cronbach's Alpha, rho_A, and Composite Reliability values were all above the minimum threshold. Cronbach's Alpha was 0.70, while the Average Variance Extracted (AVE) was above 0.50. The second-order construct, Human-Like Chatbot Interaction, had a Cronbach's Alpha of 0.904, a Composite Reliability of 0.926, and an AVE of 0.676, indicating excellent reliability and adequate convergent validity. The results of this test indicate that the Human-Like Chatbot Interaction construct is suitable for use as a higher-order construct in this research model.

4.3 Structural Model Test

Structural model testing is conducted to examine the relationship between latent constructs. The stages of structural model testing are:

Collinearity (Inner VIF) test: Collinearity (Inner VIF) testing is conducted to determine whether each predictor construct fully explains the endogenous construct in the structural model. The higher the VIF value, the greater the degree of multicollinearity, which can affect the accuracy of path coefficient estimates.

Table 6. HTMT ratio

	Human-Like Chatbot Interaction	Information Quality	Intention of Continuous Use	Service Quality	System Quality	Trust
Information Quality	0.884					
Intention of Continuous Use	0.732	0.778				
Service Quality	0.819	0.787	0.739			
System Quality	0.795	0.811	0.693	0.847		
Trust	0.806	0.821	0.785	0.805	0.799	
User Satisfaction	0.834	0.848	0.845	0.797	0.849	0.845

Table 7. Assessment of the HOC (human-like chatbot interaction)

	(β)	t-value	p-value	95% CI	Cronbach's Alpha	rho_A	Composite Reliability	AVE
Anthropomorphism	0.544	46.816	<0.001	[0.522, 0.568]	0.871	0.871	0.921	0.795
Conversational Capability	0.536	42.543	<0.001	[0.513, 0.563]	0.862	0.863	0.916	0.784
Human-Like Chatbot Interaction	-	-	-	-	0.904	0.904	0.926	0.676

According to [49], an Inner VIF value <= 5.0 indicates no multicollinearity issues among the predictor constructs. A VIF value >5.0 indicates high collinearity, which can lead to instability in structural model estimation. The results of the collinearity test (Inner VIF) shown in Table 8 show that the VIF values range from 2,412 to 3,545. All VIF values are still below the threshold of 5.0 recommended by [49], indicating that there are no multicollinearity problems in the structural model. The next stage of the structural model evaluation can be continued.

Coefficient determination test: This test is to see the results of the influence of exogenous constructs/independent variables on endogenous constructs/dependent variables. Tested based on the R-square (R^2) value, if the R^2 value >0.75 indicates a very strong model quality, if the R^2 value is between 0.50-0.75 indicates a strong model quality, if the R^2 value is between 0.25-0.50 indicates a weak model quality, and if the R^2 value <0.25 indicates a very weak model quality. The results of the R-square (R^2) test in Table 9 show that 3 endogenous constructs have values between 0.628 and 0.736, indicating that the quality of this research model strongly influences the exogenous constructs, namely Intention to Continue Use, Trust, and User Satisfaction.

Table 8. Collinearity (Inner VIF) test

Test	Intention of Continuous Use	Trust	User Satisfaction
Human Like Chatbot Interaction		3.436	3.545
Information Quality		3.316	3.520
Service Quality		3.009	3.154
System Quality		2.984	3.112
Trust	2.412		2.871
User Satisfaction	2.412		

Table 9. R-square (R^2) test

Test	R-square	R-square adjusted
Trust	0.652	0.645
User Satisfaction	0.742	0.736
Intention of Continuous Use	0.631	0.628

Effect size (f^2) test: The effect size (f^2) test is used to determine the magnitude of the influence of exogenous constructs on endogenous constructs. If the f^2 value <0.02 has a very small effect and can be ignored, the f^2 value)>=0.02 has a small effect, the f^2 value >=0.15 has a moderate effect and the f^2 value >=0.35 has a large effect [56].

In the context of PLS-SEM, [49] uses Cohen's criterion to assess the relative contribution of each exogenous construct to increasing the R^2 value of the endogenous construct. Based on Table 10, the effect size (f^2) test results indicate that the contribution of exogenous constructs to endogenous constructs ranges from 0.006 to 0.359. Based on [56] criteria adopted in PLS-SEM by [49], most relationships have small effects, while the relationship between User Satisfaction and Intention of Continuous Use shows a large effect (f^2 = 0.359). The results of this test indicate that user satisfaction is the primary driver of continued use of AI chatbots.

The Service Quality construct on User Satisfaction has an f^2 value of 0.006, indicating a very small direct contribution. This finding indicates that in the context of AI chatbots, user satisfaction is influenced not only by service quality but also by a combination of system quality, information quality, human-like chatbot interaction, and user trust. Meanwhile, the relationships between the other constructs and User Satisfaction and Trust have small effect sizes, indicating that each construct makes an additional contribution to the variance of the endogenous constructs, though not dominantly. The effect size results indicate that no single construct dominates the model. System quality, information quality, service quality, and Human-Like Chatbot Interaction simultaneously complement each other in shaping trust, user satisfaction, and Intention of Continued Use.

Predictive relevance test: The predictive relevance (Q^2) test of endogenous constructs uses the Stone-Geisser Q^2 statistic obtained through a blindfolding procedure. According to [49], a Q^2 value>0 indicates that the exogenous construct has predictive relevance to the endogenous construct. A Q^2 value >= 0.02 indicates low predictive relevance, a Q^2 value >= 0.15 indicates moderate predictive relevance, and a Q^2 value >= 0.35 indicates high predictive relevance. Based on Table 11, the predictive relevance assessment test (Q^2) shows that the Q^2 value exceeds 0.35, namely, Trust of 0.629, User Satisfaction of 0.701, and Intention of Continuous Use of 0.539. It can be concluded that the model has shown strong predictive relevance for all endogenous constructs.

Table 10. Effect size (f^2) test

Test	User Satisfaction	Intention of Continuous Use	Trust
Service Quality	0.006		0.0 48
Information Quality	0.049		0.0 61
System Quality	0.094		0..0 43
Human-Like Chatbot Interaction	0.035		0.0 32
Trust	0.074	0.085	
User Satisfaction		0.359	

Table 11. Predictive relevance (Q^2) test result

Test	Q^2predict	RMSE	MAE
Trust	0.629	0.615	0.445
User Satisfaction	0.701	0.552	0.391
Intention of Continuous Use	0.539	0.684	0.534

Predictive performance (PLSpredict) test: A predictive performance (PLSpredict) test was conducted to evaluate the out-of-sample predictive ability of the proposed model, as recommended by [57]. This evaluation compared the prediction errors of the PLS-SEM model with those of the benchmark linear model (LM). The comparison was based on the Root Mean Square Error (RMSE) value; if the RMSE value is lower than the LM, it indicates superior prediction accuracy. Table 12 shows that the TR1, TR2, TR3, US2, US3 IC1, IC2, IC3 indicators show lower RMSE values compared to LM. These results indicate that the PLS-SEM model achieves superior predictive accuracy for most indicators. However, one indicator (US1) shows a lower LM_RMSE value of 0.533 compared to the PLS-SEM_RMSE value of 0.548. If the majority of indicators show that PLS is better than LM, then the model is considered to have high predictive power [57]. The PLSpredict results show that the model still maintains good predictive ability when applied to data outside the training sample.

Path coefficient test: The path coefficient test assesses the influence of relationships between constructs using the bootstrapping method in SmartPLS4. Following the recommendations of Hair et al. (2022), the evaluation considers the standardized path coefficient (β), t-statistic, p-value, and 95% confidence interval, as well as any confidence interval that does not include zero.

Table 12. Predictive performance (PLSpredict) test results

Indicators	PLS-SEM_RMSE	LM_RMSE	Result
TR1	0.618	0.664	PLS is better
TR2	0.622	0.674	PLS is better
TR3	0.646	0.648	PLS is better
US1	0.548	0.533	LM is better
US2	0.569	0.637	PLS is better
US3	0.578	0.613	PLS is better
IC1	0.629	0.654	PLS is better
IC2	0.631	0.708	PLS is better
IC3	0.673	0.689	PLS is better

A relationship is considered statistically significant if the value (p-value <0.005) is declared 5% significant or has an effect on the construct, and vice versa if the value (p-value>0.05) is declared insignificant to the construct. Based on Table 13 of the path coefficient test, it shows that ten hypotheses are supported and one hypothesis is not supported, namely the relationship between Service Quality and User Satisfaction statistically (β = 0.068, t = 0.665, p = 0.506) because the confidence interval includes zero. Indicates that improvements in the dimensions of chatbot quality and Human-Like Chatbot Interactions contribute positively to user trust, satisfaction, and intention to continue using the service. System Quality has the strongest influence on User Satisfaction (β = 0.275, t = 3.258, p = 0.001), and User Satisfaction has the strongest influence on Continued Use Intention (β = 0.565, t = 7.752, p = 0.000). These results suggest that users will likely reuse the AI chatbot because they are satisfied with its quality.

5. Discussion and implications

5.1 Discussion

Figure 4 illustrates the optimized model.

H1: The relationship between service quality and user satisfaction, which is not statistically significant (β = 0.068, p = 0.506), is a finding that contradicts previous information system success literature. The Service Quality dimension of conventional information system success may not be the primary determinant of user satisfaction in the context of AI chatbots. This finding contradicts theories of information system success, namely the Delone and McLean model [29] and the Human Organization Technology Fit (HOT-Fit) model [30], which revealed that Service Quality is significantly related to User Satisfaction. The insignificant relationship in this study suggests that user expectations of AI chatbot services may differ from those of traditional information system services. Understanding these findings requires a broad perspective that encompasses contextual factors and the evolving nature of interactions with AI.

The comparison with previous studies differs from the predictions of the SERVQUAL model and the DeLone and McLean Information Systems Success model, which state that service quality is a determinant of user satisfaction. Similarly, the HOT-Fit model states that service quality impacts user satisfaction. Previous studies in the context of e-commerce [40], and fintech [58], found a significant relationship between Service Quality and User Satisfaction. The findings of this study indicate that User Satisfaction with AI chatbots differs from that in traditional information systems.

Table 13. Path coefficient test

Hypothesis	Relation ship	(β)	t-statistic	p-value	95% CI	Decision
H1	SQ→US	0.068	0.665	0.506	[-0.131, 0.263]	Not Supported
H2	IQ→US	0.210	2.140	0.032	[0.011, 0.390]	Supported
H3	SYQ→US	0.275	3.258	0.001	[0.106, 0.439]	Supported
H4	HLCI→US	0.179	2.624	0.009	[0.041, 0.311]	Supported
H5	SQ→ TR	0.224	2.323	0.020	[0.031, 0.408]	Supported
H6	IQ→ TR	0.266	2.729	0.006	[0.075, 0.457]	Supported
H7	SYQ→ TR	0.211	2.800	0.005	[0.066, 0.361]	Supported
H8	HLCI→ TR	0.195	2.392	0.017	[0.041, 0.363]	Supported
H9	TR→US	0.234	3.042	0.002	[0.092, 0.391]	Supported
H10	US→IC	0.565	7.752	0.000	[0.422, 0.706]	Supported
H11	TR→IC	0.274	3.744	0.000	[0.131, 0.418]	Supported

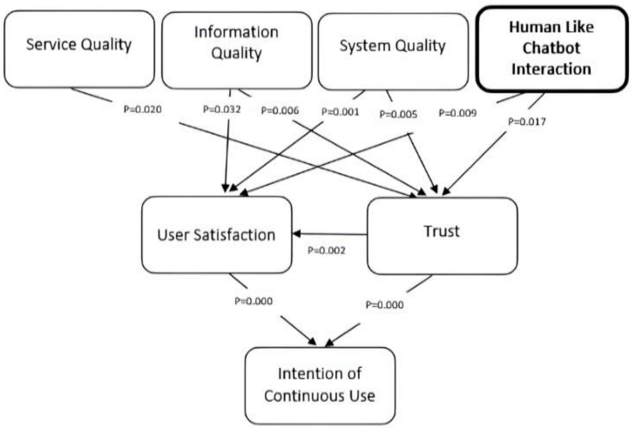

Figure 4. Optimized model

There are fundamental differences between AI chatbots and traditional information systems in context. AI chatbots are characterized by two-way interaction and communication, while traditional information systems can only interact one-way. When two-way interaction occurs, the need for AI chatbots that can think and act like human service providers, namely possessing intelligence, emotional intelligence, and social skills [59]. This is supported by a recent study showing that, in the context of AI chatbots, user satisfaction is based on the chatbot's ability to have quality conversations with a human-like communication style [60]. These results reinforce the view that user satisfaction with AI chatbot services depends more on the chatbot's intelligent interaction capabilities than on traditional service quality dimensions.

User satisfaction in an AI chatbot environment likely focuses on Information Quality, System Quality, and Human-Like Chatbot Interaction. These results are supported by the structural model test, which shows that Information Quality, System Quality, and Human-Like Chatbot Interaction are significantly related to User Satisfaction. These findings suggest that users focus more on the quality of the information they receive, good system performance, and a natural, human-like conversational experience. Trust plays an important psychological role, as users feel satisfied when the chatbot is trustworthy. This evidence suggests that user satisfaction with AI chatbot services depends more on intelligent interaction capabilities than on traditional service quality dimensions. Measurement issues may arise, even though the indicators in the Service Quality dimensions used are adapted from traditional service quality models developed for human-centered services. While these dimensions have been widely validated in conventional information systems, they may not fully capture the expectations of AI chatbot users. This is due to the development of AI technology, which allows chatbot users to increasingly expect chatbots to exhibit human-like intelligence and capabilities. Traditional service quality dimensions need to be expanded to better represent the characteristics of modern AI chatbot services. Measuring AI chatbot quality should not rely solely on conventional service quality dimensions but should also incorporate constructs oriented toward human interaction.

A paradigm shift has occurred in traditional information systems, with user satisfaction now focused on service. This contrasts with AI-based services, which focus on intelligent interaction. User satisfaction is no longer focused on the quality of service delivery but rather on information quality, system performance, and human-like interaction capabilities. The insignificant relationship between Service Quality and User Satisfaction does not imply that service quality is unimportant. This indicates that conventional service quality alone is insufficient to explain user satisfaction with AI-based chatbots. This phenomenon represents a theoretical shift from a service-oriented paradigm to a human-interactive intelligent service paradigm.

H2: Information quality is significantly related to user satisfaction.

This suggests that the higher the information quality an AI chatbot presents, the more satisfied users will be. This result is in accordance with the theory of information system success, namely the Delone and McLean model [29] measuring the quality of e-commerce and the Human Organization Technology Fit (HOT-Fit) model [30] measuring the quality of health information systems, which revealed that information quality influences User Satisfaction. Research conducted on mobile payment applications has shown that high-quality information can increase customer satisfaction [42]. Information quality, such as accuracy, completeness, up-to-dateness, clarity, and ease of understanding, can impact user satisfaction, both in the context of AI chatbots and traditional information systems.

H3: System quality is significantly related to user satisfaction.

This shows that chatbots that are easy to use, easy to access, respond quickly, and are available anytime can affect user satisfaction. These results are in accordance with the theory of information system success, namely the Delone and McLean model [29] measuring the quality of e-commerce and the Human Organization Technology Fit (HOT-Fit) model [30] measuring the quality of health information systems, which reveal that system quality has an effect on User Satisfaction. Research on mobile payment applications shows that a good system quality can increase customer satisfaction [42]. This means that AI chatbots must have a high system quality, especially those implementing modern AI technology, which requires a truly capable system. Otherwise, AI chatbots will face numerous challenges that will impact user satisfaction and trust.

H4: Human-like Chatbot Interaction is significantly related to user satisfaction.

This is a characteristic of AI chatbots that is not possessed by ordinary information systems: they can respond with greetings, communicate in a human-like speaking style and natural language, understand what is asked, and provide regular, fluent answers. This is because AI technology in chatbots enables them to exhibit human-like intelligence. This finding is supported by chatbot research in e-commerce, which indicates that friendly, empathetic chatbots can influence users [5]. In addition, in the health sector, it states that AI chatbots containing anthropomorphic aspects will appear more human, having expressions, empathy, and emotional intelligence, like characters in services provided by humans, will influence satisfaction [24], strengthened by [61], who stipulates that the quality of the conversation will influence user satisfaction. This indicates that modern AI chatbots must possess human-like intelligence, including anthropomorphism and conversational capabilities, to ensure user satisfaction with their digital services. A stiff, inhumane chatbot will negatively affect user satisfaction.

H5: Service quality is significantly related to trust.

This explains that users will trust a chatbot if it responds quickly, is secure, provides support during difficulties, and has a pleasant interface. These results align with previous research on chatbots in the fintech sector, which stated that service quality influences trust, which drives positive user experiences, and with trust, users will feel satisfied [58]. In the context of AI chatbots, users will trust the chatbot if in their experience, the services provided meet expectations and solve problems.

H6: Information quality is significantly related to trust.

This explains that the quality of information provided by chatbots, such as accurate, complete, up-to-date, clear, and easy-to-understand information, will be a factor in shaping user trust. These results are in line with e-tourism chatbot services, which state that high information quality can increase trust, customer satisfaction, and encourage reuse [62]. These results are in line with [58], who explained that chatbots with high information quality can influence trust, in line with previous research indicating that trust is the main factor in user satisfaction in the chatbot context.

H7: System quality is significantly related to trust.

This explains that chatbots that are easy to use and access, respond quickly, and are available anytime can influence user trust. Users will trust a chatbot if the chatbot guarantees system performance and stability. These results are consistent with research [63]. That system quality shapes trust in fashion service chatbots. The hotel industry also states that system quality has an influence on trust [64]. This indicates that system performance and stability are factors that support users' trust in chatbots.

H8: Human-like Chatbot interaction is significantly related to trust.

This explains that intelligent chatbots with human-like capabilities can respond with greetings, communicate in a natural, human-like speaking style, understand what is asked, and provide regular, fluent answers, which will build user trust. This is because chatbots are a substitute for human-provided services, with the hope that chatbot services can also replace them. This finding is in line with previous research in the health sector, which has shown that AI chatbots with anthropomorphic features appear more human and exhibit expressions, empathy, and emotional intelligence, like the characters in human-provided services, thereby affecting trust [24,25]. In addition, conversational capabilities in chatbots, such as natural, neat, and sequential dialogue, can increase user trust [65]. To gain user trust, a chatbot must possess human-like intelligence, anthropomorphism, and good conversational skills. A stiff and inhumane chatbot will create distrust.

H9: Trust is significantly related to user satisfaction.

Trust is significantly related to user satisfaction. These results are consistent with research on fintech chatbots, which suggests that trust plays a role in shaping satisfaction [58]. In the context of technology, trust in technology is crucial initially because it will build confidence and ultimately satisfaction [66]. This opinion strengthens the idea that users will trust a quality chatbot, which will make users feel satisfied with the chatbot service.

H10: User satisfaction is significantly related to the intention of continuous use.

These results align with the information systems success theory and the DeLone and McLean model [29] for measuring e-commerce quality, which state that user satisfaction influences intention to use. Other research on tourism chatbots also reveals that user satisfaction is a crucial factor in encouraging users to reuse chatbots [67]. In the context of AI customer service chatbots, if the chatbot meets user expectations, users will be satisfied and intend to use it again. This is supported by technical qualities such as information quality and system quality, as well as non-technical qualities such as chatbots possessing human-like intelligence, good attitudes (anthropomorphism), and good communication (conversational capability).

H11: Trust is significantly related to Intention of Continuous Use.

This explains that in the context of chatbots, the trust factor can influence users to reuse the chatbot. These results are consistent with research on telecommunications service chatbots, which states that trust influences users' intention to reuse the chatbot [68]. In banking service chatbots, trust is also stated as a major factor in chatbot reuse [69]. A chatbot that delivers consistent service and maintains a good reputation will determine whether users return to use it.

5.2 Theoretical implications

This research extends the theoretical perspective on information system success development models, specifically in the context of generative AI chatbots. This research provides theoretical contributions by adapting an extended Information system success model to support the unique characteristics of generative AI chatbots, namely, Human-Like Chatbot Interaction. Although models such as SERVQUAL [28] are considered determinants of service quality, and DeLone and McLean's Information System Success [29] emphasizes service quality, Information Quality, and System Quality as key determinants of information system success, the findings of this study indicate that technical dimensions alone are insufficient to evaluate the success of a generative AI chatbot. This is because these models were developed when information systems were still dominated by conventional applications oriented toward system functions, thus failing to consider the social interaction characteristics that are key to AI-based chatbots.

To support the unique characteristic of generative AI chatbot human-like conversations, this research extends DeLone and McLean's Information System Success model [29] by introducing a new dimension, Human-Like Chatbot Interaction (HLCI), as a quality dimension of AI chatbot conversations. HLCI is conceptualized as a higher-order construct formed by Anthropomorphism and Conversational Capability, representing a chatbot's ability to exhibit human-like behavior and communication. Many previous studies have examined these two constructs separately. The evaluation of the Higher Order Construct indicates that Anthropomorphism and Conversational Capability contribute almost equally to the HLCI construct. The perception of a chatbot as "human-like" is determined not only by anthropomorphic characteristics but also by its ability to maintain natural, responsive, and contextual conversations. HLCI can be viewed as a new conceptual construct that explains the quality of AI chatbot interactions more comprehensively than either construct alone.

Clarifying the position of Trust in the AI chatbot success model. In DeLone and McLean's model [29], the relationship between system quality and usage is primarily explained through User Satisfaction. The results show that Trust is a crucial construct linking chatbot quality with User Satisfaction and Intention of Continuous Use. All quality dimensions, Service Quality, Information Quality, System Quality, and Human-Like Chatbot Interaction, were shown to significantly influence Trust. Trust also significantly influenced both User Satisfaction and Intention to Continue Using. This reinforces the view that, in the context of AI, users evaluate not only the system's quality but also a chatbot's trustworthiness before they are satisfied and decide to continue using it.

5.3 Organizational implications

This research provides strategic implications for organizations developing or implementing Artificial Intelligence (AI) based chatbots for customer service. The research model shows that User Satisfaction is the strongest factor in increasing Intention to Continue Using, followed by Trust. The success of AI chatbot implementation is not solely measured by the initial adoption rate; efforts must focus on maintaining user satisfaction and trust to ensure continued use. For AI chatbot developers, the research findings indicate that comprehensive technical quality improvements need to be implemented through optimizing System Quality, Information Quality, and Human-Like Chatbot Interaction. Human-Like Chatbot Interaction needs to be a primary focus through developing the chatbot's ability to understand conversational context, maintain a natural dialogue flow, demonstrate empathy, and provide responses that resemble human communication. The integration of these three aspects will enhance the user experience, thereby fostering satisfaction and trust in the AI chatbot. For companies using AI chatbots as a customer service medium, the research findings indicate that chatbots should not be positioned solely as automation tools to reduce operational costs, but rather as strategic instruments to enhance the customer experience. Investment decisions in AI chatbots should focus on improving system quality, information quality, and the chatbot's interaction capabilities simultaneously, as all three aspects contribute to building user satisfaction and trust. Companies also need to build user trust by ensuring that chatbots provide consistent, transparent information and maintain the security and confidentiality of customer data. This trust will strengthen user satisfaction and ultimately increase customer intention to continue using AI chatbots as their primary service provider.

5.4 Limitations and future research

This study still has limitations, which provide opportunities for further research.

This study used a cross-sectional design, so the data represent user perceptions at a single point in time and cannot capture changes in user perceptions and behavior as user experience increases and AI chatbot capabilities develop. Given the rapid development of Large Language Model (LLM)-based chatbot technology, the relationship among chatbot quality, trust, user satisfaction, and intention to continue using chatbots could change over time. Further research is recommended using a longitudinal design to obtain stronger empirical evidence on the stability of the relationship between constructs and to evaluate the dynamics of user experience with AI chatbots across various observation periods. All constructs were measured using self-administered questionnaires by the same respondents for all constructs. This may introduce bias, although several procedural and statistical refinements were implemented to minimize it. Future research could combine questionnaire data with historical user logs for the Veronika chatbot on Telkomsel's servers to provide comprehensive support for user behavior. The respondents in this study were Indonesian users of the Veronika chatbot only. Chatbot capabilities, cultural values, communication styles, and expectations for AI-powered customer service may differ across countries and sectors, including banking, healthcare, education, public services, and e-commerce.

6. Conclusion

This study tested an AI customer service chatbot quality evaluation model using the constructs System Quality, Information Quality, Service Quality, and Human-Like Chatbot Interaction on Trust, User Satisfaction, and Intention of Continuous Use. Empirical test results indicate that System Quality, Information Quality, and Human-Like Chatbot Interaction are significantly related to Trust and User Satisfaction, while Trust and User Satisfaction are significantly related to Intention of Continuous Use. Furthermore, Service Quality was not significantly related to User Satisfaction but was significantly related to Trust. This reveals that the determining factors for AI chatbot user satisfaction are information quality, system quality, and Human-Like Chatbot Interaction, rather than service quality as in traditional information systems. This indicates a shift in the theory of the success of conventional information systems from a service-oriented approach to an intelligent, interactive one. Furthermore, the constructs Trust and User Satisfaction are significantly related to Intention of Continuous Use. The results of this validation model for evaluating the quality of AI customer service chatbots can serve as a reference in the literature on AI customer service chatbots, providing a more comprehensive framework for understanding the factors associated with trust, satisfaction, and Intention of Continuous Use in AI chatbot services.

Ethical issue

The authors are aware of and comply with best practices in publication ethics, specifically with regard to authorship (avoidance of guest authorship), dual submission, manipulation of figures, competing interests, and compliance with policies on research ethics. The authors adhere to publication requirements that the submitted work is original and has not been published elsewhere. All survey participants provided informed consent prior to participation, and the anonymity and confidentiality of all respondents were strictly maintained throughout the research process.

Data availability statement

The manuscript contains all the data. However, additional data will be provided by the corresponding author upon reasonable request.

Conflict of interest

The authors declare no potential conflict of interest.

References

[1] Smutek T, Marczuk M, Jarmul M, Jurczak E, Pliszczuk D. Chatbot to Support the Customer Service Process. Eur Res Stud J 2024;27:160–8. https://doi.org/10.35808/ersj/3396.

[2] Maher S, Bhable SG, Lahase AR, Nimbhore SS. AI and Deep Learning-driven Chatbots: A Comprehensive Analysis and Application Trends. Int. Conf. Intell. Comput. Control Syst., 2022, p. 994–8. https://doi.org/10.1109/ICICCS53718.2022.9788276.

[3] Senadheera S, Yigitcanlar T, Desouza KC, Mossberger K, Corchado J, Mehmood R, et al. Understanding Chatbot Adoption in Local Governments: A Review and Framework. J Urban Technol 2024:0–33. https://doi.org/10.1080/10630732.2023.2297665.

[4] Karimova GZ. Not in our image : rethinking anthropomorphism in expert chatbot design. AI Soc

2026;41:611–28. https://doi.org/10.1007/s00146-025-02438-z.

[5] Ding Y, Najaf M. Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce. BMC Psychol 2024;12:595. https://doi.org/10.1186/s40359-024-02083-z.

[6] Albarq A, Aljaafari AI, Mabkhot H, Alsuwaidan MS. When Chatbots Feel Human : How Anthropomorphism Shapes Consumer Satisfaction , Trust , and Loyalty in AI-Driven Brand Interactions. J Cult Anal Soc Chang 2025;10:1027–36.

[7] Ma N, Khynevych R, Hao Y, Wang Y. Effect of anthropomorphism and perceived intelligence in chatbot avatars of visual design on user experience : accounting for perceived empathy and trust. Front Comput Sci 2025. https://doi.org/10.3389/fcomp.2025.1531976.

[8] Premathilake GW, Li H. Users' responses to humanoid social robots: A social response view. Telemat Informatics 2024;91:102146. https://doi.org/10.1016/j.tele.2024.102146.

[9] Nawaz RR, Reza S, Sarwar B. Chatbot Anthropomorphism and Its Influence on Trust and Satisfaction via Perceived Value. J Soc Organ Matters 2024;3:543–68. https://doi.org/10.56976/jsom.v3i2.226.

[10] Reeves B, Nass C. The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places. Cambridge University Press.; 1996.

[11] Short J, Williams E, Christie B. The Social Psychology of Telecommunications. Wiley; 1976. ISBN: 9780608176765, 0608176761

[12] Heyselaar E. The CASA theory no longer applies to desktop computers. vol. 13. 2023. https://doi.org/10.1038/s41598-023-46527-9.

[13] Fu J, Mouakket S, Sun Y. The role of chatbots' human-like characteristics in online shopping. Electron Commer Res Appl 2023;61:101304. https://doi.org/10.1016/j.elerap.2023.101304.

[14] Huang S. Why Can Fintech Chatbots Guide Consumers to Buy Banking Products? Int J Hum Comput Interact 2025;41:12229–34. https://doi.org/10.1080/10447318.2025.2453611.

[15] Gnewuch U, Morana S, Adam MTP, Maedche A. Opposing Effects of Response Time in Human – Chatbot Interaction The Moderating Role of Prior Experience. Bus Inf Syst Eng 2022;64:773–91. https://doi.org/10.1007/s12599-022-00755-x.

[16] Følstad A, Brandtzaeg PB. Chatbots and the New World of HCI. Interactions 2017;24:38–42. https://doi.org/10.1145/3085558.

[17] Nass C, Steuer JS, Tauber ER. Computers are social actors. CHI Conf. Hum. Factors Comput. Syst., 1994, p. 72–8.

[18] Horton D, Wohl RR. Mass Communication and Para-Social Interaction. Psychiatry 1956;19:215–29. https://doi.org/10.1080/00332747.1956.11023049.

[19] Nass C, Moon Y. Machines and Mindlessness : Social Responses to Computers. J Soc 2000;56:81–103.

[20] Feine J, Gnewuch U, Morana S, Maedche A. A Taxonomy of Social Cues for Conversational Agents. J Hum Comput Stud 2019;132:138–61. https://doi.org/10.1016/j.ijhcs.2019.07.009.

[21] Jin SV, Youn S. Social Presence and Imagery Processing as Predictors of Chatbot Continuance Intention in Human-AI-Interaction Social Presence and Imagery Processing as Predictors of Chatbot Continuance. Int J Human–Computer Interact 2023;39:1874–86. https://doi.org/10.1080/10447318.2022.2129277.

[22] Cao B, Wen C, Zhang R. How Conversational Contingency Shapes Trust in Human-AI Interaction : Evidence from a Two-Wave Mediation Study. J Broadcast Electron Media 2025;69:398–417.

[23] Alwahaishi S. A Smart Interactive Behavioral Chatbot: A Theoretical Prototype. 2023 8th Int. Eng. Conf. Renew. Energy Sustain., 2023. https://doi.org/10.1109/iecres57315.2023.10209534.

[24] Schillaci CE, de Cosmo LM, Piper L, Nicotra M, Guido G. Anthropomorphic chatbots' for future healthcare services: Effects of personality, gender, and roles on source credibility, user satisfaction, and intention to use. Technol Forecast Soc Change 2024;199. https://doi.org/10.1016/j.techfore.2023.123025.

[25] Stinkeste C, Skantze G. Linguistic Anthropomorphism in Chatbots: Effects of Style, Topic, and Interaction on Users' Perceptions and Behaviors. Proc. 13th Int. Conf. Human-Agent Interact., 2025, p. 321–31. https://doi.org/10.1145/3765766.3765773.

[26] Khan R, Khan SP, Ali SA. Conversational AI. In: Rawat R, Chakrawarti RK, Sarangi SK, Vyas P, Alamanda MS, Srividya K, et al., editors. Conversational Artif. Intell., Wiley Online Library; 2024, p. 249–68. https://doi.org/10.1002/9781394200801.ch16.

[27] Lin CC, Huang AYQ, Yang SJH. A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022). Sustain 2023;15. https://doi.org/10.3390/su15054012.

[28] Parasuraman AP, Berry LL. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. J Retail 1988;64:12–40.

[29] DeLone WH, McLean ER. The DeLone and McLean model of information systems success: A ten-year update. J Manag Inf Syst 2003;19:9–30. https://doi.org/10.1080/07421222.2003.11045748.

[30] Yusof MM, Kuljis J, Papazafeiropoulou A, Stergioulas LK. An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit). Int J Med Inform 2008;77:386–98. https://doi.org/10.1016/j.ijmedinf.2007.08.011.

[31] Noor N, Rao Hill S, Troshani I. Developing a service quality scale for artificial intelligence service agents. vol. 56. 2022. https://doi.org/10.1108/EJM-09-2020-0672.

[32] Li L, Lee KY, Emokpae E, Yang SB. What makes you continuously use chatbot services? Evidence from chinese online travel agencies. Electron Mark 2021;31:575–99. https://doi.org/10.1007/s12525-020-00454-z.

[33] Pinochet LHC, de Gois FS, Pardim VI, Onusic LM. Experimental study on the effect of adopting humanized and non-humanized chatbots on the factors measure the intensity of the user's perceived trust in the Yellow September campaign. Technol Forecast Soc Change 2024;204:123414. https://doi.org/10.1016/j.techfore.2024.123414.

[34] Liu Y li, Hu B, Yan W, Lin Z. Can chatbots satisfy me? A mixed-method comparative study of satisfaction with task-oriented chatbots in mainland China and Hong Kong. Comput Human Behav 2023;143. https://doi.org/10.1016/j.chb.2023.107716.

[35] Zhang J, Chen Q, Lu J, Wang X, Liu L, Feng Y. Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism and boundary conditions. Tour Manag 2024;100:104835. https://doi.org/10.1016/j.tourman.2023.104835.

[36] Tumsekcali E, Ayyildiz E, Taskin A. Interval valued intuitionistic fuzzy AHP-WASPAS based public transportation service quality evaluation by a new extension of SERVQUAL Model: P-SERVQUAL 4.0. Expert Syst Appl 2021;186:115757. https://doi.org/10.1016/j.eswa.2021.115757.

[37] Patil K, Pramod D, Kulkarni M. Information Quality Dimensions in Generative Conversational AI for Financial Inclusion. Proc - 2024 6th Int Conf Comput Intell Commun Technol CCICT 2024 2024:64–8. https://doi.org/10.1109/CCICT62777.2024.00021.

[38] Guo Y, Dong P. Factors Influencing User Favorability of Government Chatbots on Digital Government Interaction Platforms across Different Scenarios. J Theor Appl Electron Commer Res 2024;19:818–45. https://doi.org/10.3390/jtaer19020043.

[39] Cai N, Gao S, Yan J. How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure. Humanit Soc Sci Commun 2024;11:1–11. https://doi.org/10.1057/s41599-024-03212-0.

[40] Kim J, Yum K. Enhancing Continuous Usage Intention in E-Commerce Marketplace Platforms: The Effects of Service Quality, Customer Satisfaction, and Trust. Appl Sci 2024;14. https://doi.org/10.3390/app14177617.

[41] Aoki N. An experimental study of public trust in AI chatbots in the public sector. Gov Inf Q 2020;37:101490. https://doi.org/10.1016/j.giq.2020.101490.

[42] Nilapun M, Jensuttiwetchakul T. The Effect of System Quality, Information Quality, and Service Quality on the Continued Usage of Mobile Payment Application in Thailand. ACM Int. Conf. Proceeding Ser., 2023, p. 101–8. https://doi.org/10.1145/3584871.3584886.

[43] Schillaci CE, de Cosmo LM, Piper L, Nicotra M, Guido G. Anthropomorphic chatbots' for future healthcare services: Effects of personality, gender, and roles on source credibility, user satisfaction, and intention to use. Technol Forecast Soc Change 2024;199:123025. https://doi.org/10.1016/j.techfore.2023.123025.

[44] Mayer RC, Davis JH, Schoorman FD. An Integrative Model of Organizational Trust. Acad Manag Rev 1995;20:709–34.

[45] Gao J, Promise A, Jawad C, Zhan M. The influence of artificial intelligence chatbot problem solving on customers ' continued usage intention in e-commerce platforms : an expectation-confirmation model approach. J Bus Res 2025;200:115661. https://doi.org/10.1016/j.jbusres.2025.115661.

[46] Bhattacherjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quartely 2001;25:351–70. https://doi.org/https://doi.org/10.2307/3250921.

[47] Chen Q, Gong Y, Lu Y, Tang J. Classifying and measuring the service quality of AI chatbot in frontline service. vol. 145. 2022. https://doi.org/10.1016/j.jbusres.2022.02.088.

[48] Bergmann M, Maçada ACG, de Oliveira Santini F, Rasul T. Continuance intention in financial technology: a framework and meta-analysis. Int J Bank Mark 2023;41:749–86. https://doi.org/10.1108/IJBM-04-2022-0168.

[49] Hair JF, Hult GTM, Ringle CM, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 3rd ed. Sage Publishing; 2022.

[50] Bell E, Bryman A, Harley B. Sampling in Qualitative Research. Oxford University Press; 2022. https://doi.org/10.1093/hebz/9780198869443.003.0030.

[51] Meade AW, Craig SB. Identifying careless responses in survey data. Psychol Methods 2012;17:437–55. https://doi.org/10.1037/a0028085.

[52] Kock N. Common Method Bias in PLS-SEM: A Full Collinearity Assessment Approach. Int J e-Collaboration 2015;11:1–10. https://doi.org/10.4018/ijec.2015100101.

[53] Jais R, Ngah AH. The moderating role of government support in chatbot adoption intentions among Malaysian government agencies. Transform Gov People, Process Policy 2024;18. https://doi.org/10.1108/TG-02-2023-0026.

[54] Li L, Lee KY, Emokpae E, Yang SB. What makes you continuously use chatbot services? Evidence from chinese online travel agencies. Electron Mark 2021;31:575–99. https://doi.org/10.1007/s12525-020-00454-z.

[55] Liu W, Jiang M, Li W, Mou J. How does the anthropomorphism of AI chatbots facilitate users' reuse intention in online health consultation services? The moderating role of disease severity. Technol Forecast Soc Change 2024;203:123407. https://doi.org/10.1016/j.techfore.2024.123407.

[56] Cohen J. Statistical Power Analysis for the Science. 2nd ed. NJ: Lawrence Erlbaum Associates; 1988. https://doi.org/10.4324/9780203771587

[57] Shmueli G, Hair JF, Ting H, Ringle CM. Predictive model assessment in PLS-SEM : guidelines for using PLSpredict. Eur J Mark 2019;53:2322–47. https://doi.org/10.1108/EJM-02-2019-0189.

[58] Goh TS, Rohim A. AI Chatbots Driving Fintech User Satisfaction via Trust and Service Quality. 2025 4th

Int. Conf. Creat. Commun. Innov. Technol., IEEE; 2025, p. 1–7. https://doi.org/10.1109/ICCIT65724.2025.11167781.

[59] Huang MH, Rust RT. A strategic framework for artificial intelligence in marketing. J Acad Mark Sci 2021;49:30–50. https://doi.org/10.1007/s11747-020-00749-9.

[60] Al-Shafei M. Navigating Human-Chatbot Interactions: An Investigation into Factors Influencing User Satisfaction and Engagement. Int J Hum Comput Interact 2024;0:1–18. https://doi.org/10.1080/10447318.2023.2301252.

[61] Hsu CL, Lin JCC. Understanding the user satisfaction and loyalty of customer service chatbots. J Retail Consum Serv 2023;71. https://doi.org/10.1016/j.jretconser.2022.103211.

[62] Masri NW, You JJ, Ruangkanjanases A, Chen SC, Pan CI. Assessing the effects of information system quality and relationship quality on continuance intention in e-tourism. Int J Environ Res Public Health 2020;17. https://doi.org/10.3390/ijerph17010174.

[63] Shahzad MF, Xu S, An X, Javed I. Assessing the impact of AI-chatbot service quality on user e-brand loyalty through chatbot user trust, experience and electronic word of mouth. J Retail Consum Serv 2024;79:103867. https://doi.org/10.1016/j.jretconser.2024.103867.

[64] Nguyen VT. The impact of AI chatbots on customer trust: an empirical investigation in the hotel industry. Consum Behav Tour Hosp 2023;18:293–305. https://doi.org/10.1108/CBTH-06-2022-0131.

[65] Chen J-H, Guo F, Ren Z, Li M, Ham J. Effects of Anthropomorphic Design Cues of Chatbots on Users' Perception and Visual Behaviors. Int J Hum Comput Interact 2023:1–19. https://doi.org/10.1080/10447318.2023.2193514.

[66] Wut TM, Lee SW, Xu J, Kwok MLJ. Do Trusting Belief and Social Presence Matter? Service Satisfaction in Using AI Chatbots: Necessary Condition Analysis and Importance-Performance Map Analysis. Informatics (Basel) 2025;12:91. https://doi.org/10.3390/informatics12030091.

[67] Orden-Mejía M, Carvache-Franco M, Huertas A, Carvache-Franco O, Carvache-Franco W. The Role of AI-Based Destination Chatbots in Satisfaction, Continued Usage Intention, and Visit Intention: A Study from Quito, Ecuador. Int J Hum Comput Interact 2024;41:9384–99. https://doi.org/10.1080/10447318.2024.2425882.

[68] Alhasasneh S, Shishan F, Balushi W Al. Factors influencing continuous intention to use chatbots and their socio-economic development implications: evidence from Jordan's telecommunications industry. J Bus Socio-Economic Dev 2025. https://doi.org/10.1108/jbsed-07-2025-0295.

[69] Nguyen DM, Chiu YTH, Le HD. Determinants of continuance intention towards banks' chatbot services in vietnam: A necessity for sustainable development. Sustain 2021;13. https://doi.org/10.3390/su13147625.