Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10, 1416-1436
2025
Publisher: Learning Gate
DOI: 10.55214/2576-8484.v9i10.10674

AI chatbot quality in customer service: Extending measurement models with conversational capability

Warjiyono[1*], Hadiyanto[2], Dinar Mutiara Kusumo Nugraheni[3]
[1]Doctoral Program of Information System, Postgraduate School, Universitas Diponegoro, Semarang, Indonesia; warjiyono@students.undip.ac.id (W.).
[2]Department of Chemical Engineering, Faculty of Engineering, Universitas Diponegoro, Semarang, Indonesia; hadiyanto@live.undip.ac.id (H.).
[3]Department of Informatics, Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia; dinar.mutiara@live.undip.ac.id (D.M.K.N.).

Abstract: The development of artificial intelligence (AI) is driving the use of chatbots to improve customer service. However, measuring the quality of chatbots remains a crucial issue despite the importance of customer satisfaction. This research aims to develop a model for measuring the quality of AI chatbot customer service by integrating technical dimensions, conversational capability, and anthropomorphism to align with advancements in modern AI chatbots. This research adopts established models such as SERVQUAL, DeLone and McLean, HOT fit, and AICSQ. The quantitative approach uses Partial Least Squares Structural Equation Modeling (PLS-SEM). Data was collected in two stages: a pretest with 83 respondents and a nomological test with 213 respondents using the Telkomsel Veronika chatbot. The test results show that only the information quality and system quality variables significantly directly influence satisfaction, and only the conversational capability variable directly influences satisfaction and trust, which strengthens the intention to continue using the chatbot. This research contributes to the development of a chatbot evaluation model by highlighting the importance of technical and conversational capability dimensions in shaping user experience. Further research could expand the context to various service sectors, as well as consider local cultural and linguistic factors.
Keywords: *AI chatbot, Conversational capability, Customer service, Intention of continuous use, Trust, User satisfaction.*

1. Introduction

Currently, the use of artificial intelligence (AI) worldwide continues to increase. Chatbots have become very popular in the private sector, such as banking, media, and tourism [1]. Chatbots have been widely implemented by companies for customer service [2] to provide information and serve customers because they can reduce employee workload [3], service costs, service waiting times, and the ability to create openness and transparency of information [4]. As many as 55% of companies use chatbots to support customer service. Chatbots are most frequently used for sales (41%), customer support (37%), and marketing (17%) [5].

Chatbot AI leverages advanced technologies such as Natural Language Processing (NLP) and Machine Learning (ML) to provide personalized, real-time communication, thereby improving service delivery and customer satisfaction. As AI technology advances, chatbots can now perform more complex tasks, such as information and document retrieval [4]. Chatbots are predicted to become a necessity for companies in customer service, as they are more proactive, respond quickly to customer questions, and are available 24/7 for customer satisfaction [6]. For companies, using chatbots can save costs and provide better and faster service [7].

History: Received: 12 August 2025; Revised: 28 August 2025; Accepted: 2 September 2025; Published: 23 October 2025
* Correspondence: warjiyono@students.undip.ac.id

Although chatbots can offer automated solutions, the quality of the interactions they provide remains a crucial issue in determining customer satisfaction levels [8]. Measuring the quality of chatbots is not only important to ensure they can provide fast and accurate responses, but also to improve the overall customer experience. A high-quality chatbot is one that can effectively understand and respond to customer needs, significantly reducing service time and potentially increasing user engagement, loyalty, and satisfaction [9]. Additionally, the quality of communication between chatbots and customers, such as anthropomorphism and communication style, plays a significant role in determining user satisfaction and driving greater consumer engagement [10]. On the other hand, poor chatbot quality can negatively impact customer experience and even decrease customer trust in the company's services. Chatbots that frequently fail to meet user expectations, such as failing to understand and resolve complex questions, can lead to negative customer emotions [11]. Additionally, when chatbots do not provide good responses, a lack of empathy can lead to customer dissatisfaction with the service [12].

Therefore, effective chatbot quality measurement can help identify existing shortcomings in a company's chatbot system, as well as provide solutions for improvements that can enhance the overall customer experience. Measuring chatbot quality includes analyzing how chatbots can create emotional connections with users. Interactions that evoke empathy and understanding can strengthen the relationship between customers and companies, as well as increase customer satisfaction and loyalty [9].

Given the widespread use of chatbots across various fields, many researchers are also studying the scale of service quality in the business sector [13]. However, the issue of service quality provided by chatbots still poses a challenge, particularly in terms of customer satisfaction [8]. One of the main issues is how to measure the quality of chatbots, as the standards for measuring chatbot quality are not entirely clear, as evidenced by the diverse approaches and frameworks proposed in recent research. As AI chatbots continue to advance, ongoing research will be needed to align service quality metrics with new capabilities and user expectations [14]. An effective chatbot should be able to provide relevant information, manage interactions naturally, and handle customer requests quickly and accurately [9]. On the other hand, factors such as understanding the context of the conversation, response speed, and its ability to provide adequate solutions also play a role in assessing the quality of a chatbot. According to Song et al. [15] it is necessary to analyze chatbot user experience to achieve higher levels of trust.

The leading theoretical framework used for measuring and evaluating the quality of both offline and application-based services is the SERVQUAL model [16]. Many researchers have used SERVQUAL to measure service quality, such as in public health services, educational services, public transportation services, tourism services, government employment services, and urban services. However, the dimensions and scale of service quality become a challenge when implemented within the framework of AI chatbot services. First, the five SERVQUAL dimensions, tangibles, reliability, responsiveness, assurance, and empathy, do not fully represent the characteristics of AI chatbot services. Second, traditional human-based services differ from AI chatbot services, which are supported by advanced technology that allows AI chatbots to have human-like capabilities such as answering questions and describing product features [3], expressing emotions [11], and thinking intelligently [17].

Additionally, the HOT-FIT (Human, Organization, and Technology Fit) method [18] is very important because it covers three aspects that influence technology acceptance and success: human, organizational, and technological aspects. The hot-fit model is quite adaptable because it can be used to evaluate the effectiveness and efficiency of information systems, especially in the public sector [19] and to evaluate hospital Electronic Medical Record (EMR) systems [20]. However, the dimensions and measurement scale of HOT-FIT, when applied to measure the quality of AI chatbots, pose a challenge because they do not explicitly include transparency, data security, and ethics [19]. In fact, in customer service, chatbots handle personal data, provide accurate responses, and offer conversational interactions.

Previous models and theories in the development of this chatbot quality measurement model are important because, first, the multidimensional characteristics of chatbots refer to their ability to handle various dimensions or aspects of interaction and service. By synthesizing established models, the quality assessment of chatbots can be more comprehensive and remain relevant to current developments. Second, each quality measurement model has its own limitations. By combining several models, the shortcomings of one model can be overcome with the strengths of others, providing more accurate and balanced measurement results [21]. Third, for applications operating in different environments, quality measurement needs to be adjusted to account for various conditions and requirements [22].

This study aims to develop a model to measure the quality of AI chatbot services within the scope of customer service, serving as a framework for evaluating and improving the quality of AI chatbot services in this domain. To expand this research, two research questions are posed: first, what are the comprehensive dimensions for measuring the quality of AI chatbot services within the scope of customer service? Second, how can the quality of AI chatbot services be effectively measured within this scope?

To develop a measurement model for AI chatbot customer service quality, a quantitative method is used through outer model testing, namely convergent validity testing, discriminant validity testing, reliability testing, and inner model testing, namely path coefficient testing and specific indirect effect testing. This AI chatbot customer service quality measurement model goes through three rounds of data collection and statistical testing: the pretest stage, the instrument strengthening stage, and the nomological testing stage.

The results of this study contribute to the study of measuring the quality of AI chatbot services, particularly in customer service. Theoretically, this study aims to develop dimensions that comprehensively meet the context of advanced AI chatbot services. Empirically, this research constructs an AI chatbot service quality measurement model based on user perceptions of satisfaction with AI chatbot customer service usage. Practically, it produces a comprehensive model that can be used to evaluate and assess chatbot quality in a holistic customer service context.

2. Theoretical Background

2.1. Chatbot AI in Customer Service

The development of chatbots in customer service has undergone a significant transformation with advancements in artificial intelligence (AI) technology. Initially, chatbots were used to automatically answer basic questions. However, they can now handle more complex tasks, such as problem-solving, service personalization, and more natural interactions with customers.

The development of AI technology has enabled chatbots to use natural language processing (NLP) to answer questions and describe product features [3] in a manner similar to humans providing service, as they possess the ability to express emotions [11] and think intelligently [17]. Chatbots are generally integrated with websites and online social networks [4]. Chatbot AI can provide users with the information they need by utilizing existing data from various sources, such as documents, websites, operational systems, and social media [1]. Chatbots offer several advantages, such as the ability to answer questions and provide solutions faster and more efficiently than human-to-human interaction. Users can ask questions anytime, anywhere, and receive quick responses, allowing for 24-hour communication without time constraints [23]. They are user-friendly, make users feel safe, and simplify daily life.

Overall, the implementation of AI chatbots in customer service has brought positive changes to customer relationship management by providing real-time responses and improving service efficiency. However, challenges such as technological anxiety and the need for transparency still need to be addressed to ensure optimal adoption and effectiveness [24]. With rapid changes in technology and social structures, companies are increasingly striving to build a competitive advantage through improved service quality, which is a key factor influencing user satisfaction and behavior.

2.1.1. Quality of AI Chatbot Service Customer

Services generally involve two subjects: the service provider (organization) and the service user (customer). The service provider will strive to the best of their ability to provide the best service to the customer to ensure customer satisfaction. Advances in artificial intelligence (AI) technology are driving the development of AI chatbots to enhance customer service, making them the latest innovative service. This allows AI chatbots to be anthropomorphic, meaning they possess emotional intelligence and characteristics similar to human-provided services, but can surpass humans in certain aspects of intelligence, such as data and information storage, computing power, and learning ability.

More and more research is exploring the dimensions and scales of website-based service quality, such as WEBQUAL [25], EGSQUAL [26], and it is developing in various fields such as retail services and in various other industries and sectors [16]. Many researchers are using the dimensions of the SERVQUAL model [16] to measure service quality, such as public health services [27], education services [28], public transportation services [29], tourism services [30], government employment services [31], and urban services [32].

Additionally, the DeLone and McLean [33] is widely used to assess the quality of information systems in areas such as healthcare [34], fintech [35], accounting [36], m-banking [37], e-learning [38], and tourism [30]. Furthermore, the HOT-fit model [18] is also widely used to evaluate the effectiveness and efficiency of information systems, especially in the public sector [19] and to evaluate hospital Electronic Medical Record (EMR) systems [20].

However, these dimensions for measuring website-based service quality are not comprehensive enough to be applied to AI chatbots. This is because AI chatbots have advanced features that can perform more complex tasks, such as information and document retrieval [4]. Additionally, AI chatbots adopt anthropomorphic characteristics, meaning they can possess emotional intelligence, such as using emoticons to represent attitudes and emotional expressions [39]. Furthermore, conversations with AI chatbots are designed to be increasingly interactive [40] with communication characteristics similar to the service provided by humans, featuring a natural language style and familiar interactions [41].

Thus, the emergence of these new dimensions underscores the need for updating the service quality evaluation model to better align with the context of AI chatbots in customer service, particularly the unique aspects of AI technology that integrate anthropomorphism and communication capabilities into the user interaction experience.

2.1.2. Measuring the Quality of AI Chatbot Services

The AI chatbot quality measurement model developed by Chen et al. [42] called AICSQ (Artificial Intelligence Customer Service Quality) has seven constructs: semantic understanding, close human-AI collaboration, human-like, continuous improvement, personalization, cultural adoption, and efficiency. The findings support that all seven dimensions of AICSQ positively influence user satisfaction, perceived value, and intention to continue using AI chatbots. The AICSQ model has been established for the evaluation of AI chatbots. The AICSQ model already includes a human-like dimension, but it does not yet emphasize the technical aspects of conversation. Therefore, "human-like" in the AICSQ study as a single variable needs to be broken down into a more operational form through anthropomorphism and conversation capability.
Future research directions will expand human-like capabilities into two important perspectives: anthropomorphism and conversational capability. Anthropomorphism emphasizes the extent to which chatbots exhibit human-like attributes such as persona, communication style, and empathy, while conversational capability reflects the technical ability of chatbots to maintain natural, coherent, and interactive conversations just like humans. This approach is expected to provide a deeper understanding of how human-like characteristics in chatbots contribute to improved service quality and customer satisfaction. Additionally, based on the recommendations for future research, it is requested that trust factors influencing customer intention with AI chatbots be included, as well as the development of AICSQ for various industries and chatbot types [42].

3. Methodology

The development and measurement of the AI chatbot quality construct in this study integrates several methodological frameworks to be more systematic and comprehensive. First, Churchill's purification paradigm [43] served as the initial basis for systematic steps in developing and refining construct measurement instruments to achieve high validity and reliability. Second, the approach

Anderson and Gerbing [44] use is a two-step SEM approach that separates the testing of the measurement model and the structural model. Third, the approach Hair et al. [45] provides technical guidance for evaluating PLS-SEM models. The integration of these three approaches results in a comprehensive seven-step procedure capable of producing valid and reliable instruments, as well as a model that is fit for measuring the quality of AI chatbots in customer service. Here are the seven steps of research:

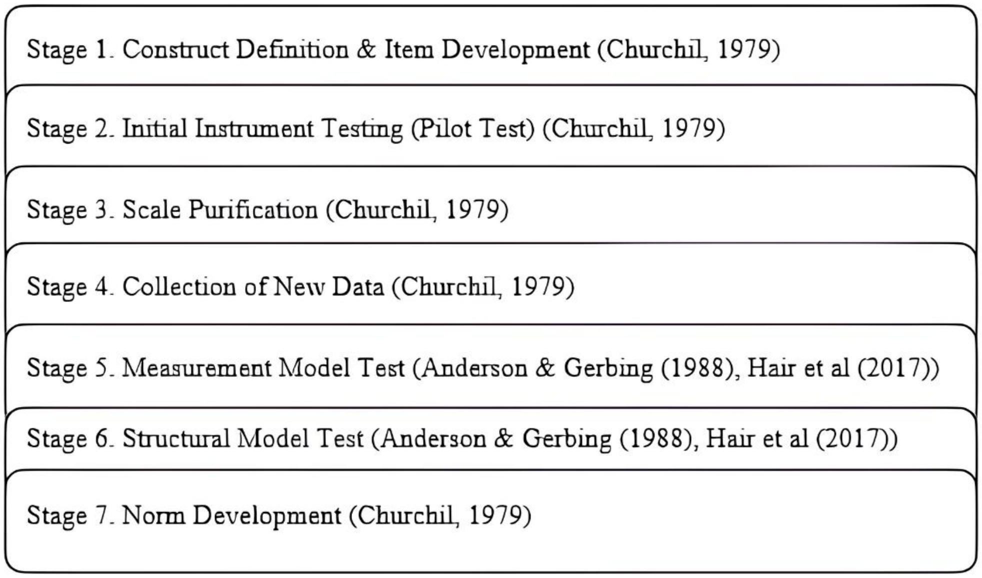

Figure 1.
Research Stages.
Source: Churchill Jr [43]; Anderson and Gerbing [44] and Hair et al. [45].

4. Results

4.1. Construct Definition & Item Development

The first step is to define the construct, which was done based on a literature review of scales relevant to service quality. The selection of dimensions for measuring the quality of AI chatbots in this study is based on a synthesis of several information system and service quality evaluation models, such as SERVQUAL Parasuraman, et al. [16], DeLone and McLean [33], HOT-fit [18] and AICSQ [42] and is further supported by contributions from DeLone and McLean [33] and is further supported by contributions from DeLone and McLean [33] model and the Hot-fit model, which underscores the importance of chatbot interface design elements, response speed, assurance, empathy, and technical support in influencing users' perceived service quality. This dimension holds particular relevance in the context of AI chatbots, as the quality of digital service is always evaluated based on how chatbots provide satisfying responses, support, and interactions.

System Quality is adapted from the DeLone & McLean model and HOT-fit, which emphasizes technical quality, such as the technical quality of the chatbot, including ease of use, response speed, system reliability, accessibility, access speed, and security. In AI chatbots, this factor becomes crucial because system stability and ease of interaction determine the success of the user experience.

Information Quality is adapted from the DeLone & McLean model and HOT-fit, which emphasize the quality of information provided by chatbots, such as information accuracy, relevance, up-to-date information, and clarity. In the context of chatbots, information quality determines whether users feel helped and trust the answers provided by the chatbot regarding the truth and novelty of the information.

Anthropomorphism is adopted from the latest AICSQ (Artificial Intelligence Customer Service Quality) model by Chen et al. [42] called Human-like, which emphasizes the anthropomorphic nature of AI, namely the ability of chatbots to exhibit human-like qualities such as emotional intelligence, personal closeness, and emotional expression. This is important because anthropomorphism has been proven to increase psychological closeness and user satisfaction [46] and can increase user trust [47]. These findings collectively highlight the significant influence of anthropomorphism on user satisfaction across various technological interfaces and applications.

Conversational Capability is adapted from the Conversational Flow dimension created by Hmoud et al. [48], who successfully developed and validated a framework in the form of a rubric to assess the effectiveness of AI chatbots. Additionally, based on Human-Computer Interaction (HCI), which is recognized for its important role in developing user-friendly and effective computing systems, particularly AI-based chatbots [49]. HCI has evolved to prioritize user experience, focusing on intuitive and natural interactions. Chatbots, as conversational interfaces, should facilitate user experience, emphasizing intuitive interactions and natural dialogue [50]. The importance of conversation quality in chatbots was also highlighted by Silva and Canedo [51], who stated that the naturalness and clarity of a chatbot are key conversational attributes that can influence user satisfaction. As AI technology, particularly NLP, continues to advance in interpreting and generating human language, techniques such as Natural Language Understanding and Natural Language Generation are crucial for creating contextually coherent and relevant responses[52].

Table 1.
Definitions of AI Chatbot Dimensions.

Dimensions	Description	Reference
Service Quality	Evaluating service aspects provided by chatbots, such as tangibles, responsiveness, assurance, empathy, and technical support.	Parasuraman et al. [16], Yusof et al. [18], DeLone and McLean [33], Chen et al. [42], Møller et al. [53] and Ashfaq et al. [54]
System Quality	Assessing the technical quality level of chatbots, including ease of use, reliability, response time, security, accessibility, and availability.	Yusof et al. [18] and DeLone and McLean [33]
Information Quality	Assessing the quality of information provided by chatbots, such as accuracy, completeness, relevance, timeliness, and ease of understanding.	Yusof et al. [18], DeLone and McLean [33], Møller et al. [53] and Ashfaq et al. [54]
Anthropomorphism	Assessing the chatbot's ability to exhibit human-like qualities, such as human-like social cues, human-like personality, and emotional capability.	Chen et al. [42], Møller et al. [53], Noor et al. [55] and Zhang et al. [56]
Conversational Capability	Assessing chatbot capabilities in managing conversations, such as naturalness, language appropriateness, and engagement.	Chen et al. [42], Møller et al. [53], Hsu and Lin [57] and Li et al. [58]
User Satisfaction	Measuring user satisfaction with the experience of interacting with the chatbot, both emotionally, cognitively, and with the overall service.	DeLone and McLean [33] and Chen et al. [42]
Intention of Continuous Use	Users' intention to continue using AI chatbots in the future	DeLone and McLean [33], Chen et al. [42] and Kim and Yum [59]
Trust	User trust in AI chatbots	DeLone and McLean [33], Ciechanowski et al. [47] and Noor et al. [55]

ISSN: 2576-8484
Vol. 9, No. 10: 1416-1436, 2025
DOI: 10.55214/2576-8484.v9i10.10674

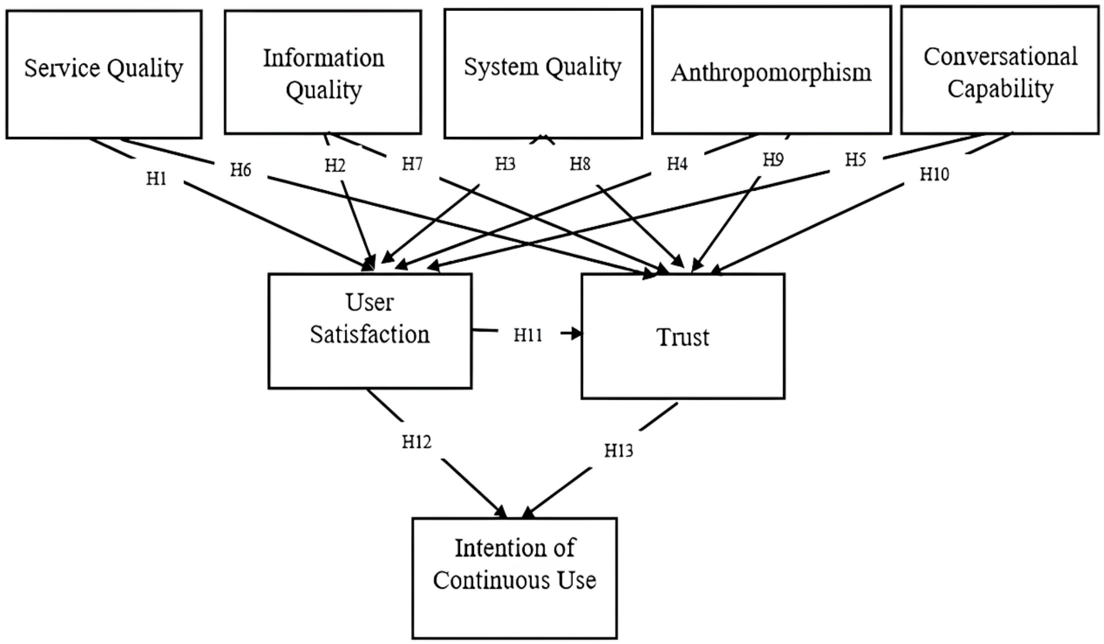

Figure 2.
Proposed Research Model.

Figure 2 is a conceptual model illustrating the hypothesis of the relationship between the main determinants of AI chatbot customer service quality. Variables such as service quality, information quality, system quality, anthropomorphism, and conversational capability are expected to temporarily influence user satisfaction and trust in AI chatbots. In turn, user satisfaction and trust are expected to affect users' intention to continue using the chatbot.

4.2. Construct Definition & Item Development

Initial data collection in this study was conducted by distributing questionnaires using a 1–5 Likert scale, where scale 1 indicates "strongly disagree" and scale 5 indicates "strongly agree." The questionnaire instrument was developed based on the indicators that had been established in the previous stage, with the aim of testing the extent to which the statements in the questionnaire could be clearly understood by the respondents and did not lead to multiple interpretations. The respondents in this initial study were users of Telkomsel's Veronika chatbot, selected for their relevance to the research context regarding the quality of AI chatbots in customer service. The number of respondents participating in this initial stage was 83. The questionnaire distribution method was carried out in two ways: online and offline. Online distribution utilized a Google Form link, which was shared through Telkomsel user communication groups or networks. Meanwhile, offline distribution involved delivering printed questionnaires to users who had directly interacted with the Veronika chatbot. This initial data collection phase was not intended to test the research hypothesis but rather focused on evaluating whether the questionnaire instrument, particularly the developed indicator items, could be well understood by the respondents. Thus, the results of this initial data collection will serve as the basis for refining the instrument before proceeding to further testing with a larger number of respondents.

4.3. Scale Purification

The scale purification stage is carried out to ensure that the research instrument used has an adequate level of reliability and validity before being used for model testing. First, reliability testing was conducted using Cronbach's Alpha and Composite Reliability (CR) values; values above 0.70 indicate that the statement items are consistent across the measured variables [60]. Second, convergent validity is considered fulfilled if the AVE value is above 0.50, which confirms that the indicators within a construct are correlated with each other.

Table 2.
Scale Purification.

	Cronbach's alpha	Composite reliability (rho_c)	Average variance extracted (AVE)
System Quality	0.885	0.913	0.636
Information Quality	0.882	0.914	0.680
Service Quality	0.889	0.919	0.693
Anthropomorphism	0.781	0.872	0.694
Conversational Capability	0.739	0.851	0.656
User Satisfaction	0.843	0.906	0.762
Trust	0.705	0.829	0.617
Intention of Continuous Use	0.776	0.870	0.690

Based on Table 2, the data processing results show that all constructs meet the criteria for reliability and convergent validity. The System Quality construct obtained a Cronbach's Alpha value of 0.885, a Composite Reliability (CR) of 0.913, and an Average Variance Extracted (AVE) of 0.636. The Information Quality construct had a Cronbach's Alpha of 0.882, CR of 0.914, and AVE of 0.680. Both constructs demonstrate excellent internal consistency and adequate convergent validity. For the Service Quality construct, Cronbach's Alpha was 0.889, CR was 0.919, and AVE was 0.693, indicating very strong reliability. The Anthropomorphism construct also meets the criteria with a Cronbach's Alpha of 0.781, CR of 0.872, and AVE of 0.694. The Conversational Capability construct has a Cronbach's Alpha of 0.739, CR of 0.851, and AVE of 0.656, which, although relatively lower than the other constructs, still exceeds the recommended threshold. Furthermore, the User Satisfaction construct demonstrates high internal consistency with a Cronbach's Alpha of 0.843, CR of 0.906, and AVE of 0.762. The Trust construct is also deemed reliable with a Cronbach's Alpha of 0.705, CR of 0.829, and AVE of 0.617. The Intention of Continuous Use construct achieves a Cronbach's Alpha of 0.776, CR of 0.870, and AVE of 0.690, indicating that the instrument has good reliability and strong convergent validity. Overall, these results indicate that all constructs used in the study have met the criteria for reliability and convergent validity. Therefore, the research instrument can be declared suitable for use in the subsequent structural model analysis stage.

4.4. New Data Collection

After the scale purification stage resulted in a valid and reliable research instrument, the next step was to collect new data from a larger sample of 213 respondents. This number was chosen to meet the needs of quantitative analysis, particularly in the Partial Least Squares–Structural Equation Modeling (PLS-SEM) approach, which requires a sufficiently large sample size for more reliable and representative analysis results [61].

The questionnaire was distributed online through the Google Forms platform. The data obtained at this stage was no longer solely to test respondents' understanding of the statement items but was used directly for testing the research model. Using this new data, a PLS-SEM analysis was conducted to test the relationships between latent variables, assess the quality of the measurement model and the structural model, and test the formulated hypotheses. With new data collection involving a larger number of respondents, this study is expected to provide more methodologically robust testing results and support more valid conclusions regarding the quality of AI chatbots in user experience-based customer service.

4.5. Test of the Measurement Model (Outer Model)

After a sufficient amount of data has been collected, the next step is to assess the validity and reliability of the research instrument. This testing aims to ensure that the constructs used are truly well-measured and can accurately represent the latent variables.

First, reliability testing was conducted: Cronbach's Alpha and Composite Reliability (CR ≥ 0.70). A construct is considered reliable if its CR value is greater than 0.70 [60].

Table 3.
Reliably Testing: Cronbach's Alpha and Composite Reliability.

	Cronbach's alpha	Composite reliability (rho_c)	Average variance extracted (AVE)
System Quality	0.913	0.913	0.933
Information Quality	0.913	0.915	0.935
Service Quality	0.908	0.912	0.931
Anthropomorphism	0.860	0.866	0.915
Conversational Capability	0.875	0.877	0.923
User Satisfaction	0.929	0.929	0.955
Trust	0.889	0.890	0.931
Intention of Continuous Use	0.932	0.932	0.957

Table 3, which illustrates all the constructs in this study, shows excellent criteria. The System Quality construct has a Cronbach's Alpha of 0.913, CR of 0.913, and AVE of 0.933. This indicates very high internal consistency and the indicator's ability to explain construct variance exceeding the minimum threshold. The Information Quality construct obtained a Cronbach's Alpha of 0.913, CR of 0.915, and AVE of 0.935, which also indicates strong reliability and convergent validity. Similarly, the Service Quality construct recorded a Cronbach's Alpha of 0.908, CR of 0.912, and AVE of 0.931. The Anthropomorphism construct had values of Cronbach's Alpha 0.860, CR 0.866, and AVE 0.915, while Conversational Capability obtained Cronbach's Alpha 0.875, CR 0.877, and AVE 0.923. The User Satisfaction construct demonstrated Cronbach's Alpha of 0.929, CR of 0.929, and AVE of 0.955. Furthermore, the Trust construct also meets the criteria with Cronbach's Alpha 0.889, CR 0.890, and AVE 0.931. Finally, the Intention of Continuous Use construct demonstrates excellent reliability and convergent validity with a Cronbach's Alpha of 0.932, CR of 0.932, and AVE of 0.957. Overall, the test results indicate that all constructs used in the study have met the requirements for reliability and convergent validity with very satisfactory values. This indicates that the research instrument is truly capable of measuring the intended constructs, making it suitable for use in structural model testing at the next stage.

Second, convergent validity testing was conducted using the Convergent Validity: Outer Loading (≥0.70) value [62]. The results of the outer loading test show that all indicators for each construct have values above 0.70, indicating that these indicators effectively reflect the measured constructs. For example, Anthropomorphism (3 indicators) has loadings between 0.852 and 0.909, and Conversational Capability (3 indicators) has loadings between 0.891 and 0.902. Similarly, indicators for other constructs such as Information Quality (5 indicators) range from 0.828 to 0.889, Service Quality (5 indicators) from 0.806 to 0.890, System Quality (6 indicators) from 0.733 to 0.876, Trust (3 indicators) from 0.891 to 0.913, User Satisfaction (3 indicators) from 0.923 to 0.947, and Intention of Continuous Use (3 indicators) from 0.935 to 0.940. All these values exceed the recommended minimum threshold of 0.70, thus satisfying the criteria for convergent validity. Therefore, it can be concluded that the indicators used in this research model are valid and appropriate for measuring the relevant constructs.

Third, discriminant validity testing was conducted using the Fornell-Larcker criterion, cross-loading, and HTMT (<0.90) to assess whether the constructs were truly distinct from other constructs. An HTMT value less than 0.90 indicates that the tested constructs have good discrimination, so there is no overlap between the latent variables [63]. A construct is said to have good discriminant validity if the HTMT value is below the threshold of 0.90 [63], although for exploratory research, values up to 0.95 are still acceptable [64]. The HTMT test indicates that most of the HTMT values between constructs are below 0.90, thus meeting the criteria for discriminant validity.

However, there are several pairs of constructs that are close to the threshold, such as between Anthropomorphism and Conversational Capability (0.938), Conversational Capability and Information Quality (0.941), Conversational Capability and Service Quality (0.943), TR and Intention of Continuous Use (0.911), User Satisfaction and Intention of Continuous Use (0.948), Information Quality and Service Quality (0.933), and Information Quality and System Quality (0.932). This value is still acceptable because it falls within the commonly used tolerance range of 0.95 for exploratory research [64]. Thus, the constructs in this study are declared to have adequate discriminant validity, making the outer model suitable for testing the structural model.

4.6. Test of the Structural Model (Inner Model)

This test aims to evaluate the relationships between latent constructs in the research model and assess the model's ability to explain endogenous variables. Some criteria used in evaluating the inner model include: first, the coefficient of determination (R^2); second, effect size (f^2) and predictive relevance (Q^2); third, SRMR (Standardized Root Mean Square Residual); and fourth, bootstrapping. The coefficient of determination (R^2) is used to assess the ability of exogenous variables to explain endogenous variables.

First, the coefficient of determination (R^2) is used to assess the ability of exogenous variables to explain endogenous variables.

Table 4.
Determinantion Coefficient Test (R^2).

	R-square	R-square adjusted
Intention of Continuous Use	0.835	0.834
Trust	0.733	0.725
User Satisfaction	0.763	0.757

Table 4 shows that the Intention of Continuous Use construct has an R-squared value of 0.835 (adjusted 0.834). This means that the exogenous variables in the model are able to explain 83.5% of the variability in the Intention of Continuous Use, with the remaining 16.5% influenced by factors outside the research model. This value falls into the very strong category according to the criteria [65]. The Trust construct has an R-squared value of 0.733 (adjusted 0.725). This indicates that 73.3% of the variation in Trust can be explained by the exogenous constructs influencing it, with the remaining 26.7% attributed to other factors outside the model. According to classification standards, this value falls into the strong category. The User Satisfaction construct has an R-squared value of 0.763 (adjusted 0.757), meaning that 76.3% of the variation in User Satisfaction is explained by the exogenous constructs within the model, while 23.7% is due to external factors. Overall, the R-squared values for the three endogenous constructs are in the strong category, indicating that the structural model developed in this study possesses high predictive capability.

Second, the effect size (f^2) is used to assess the predictive power of the model. Referring to the criteria [66] an f^2 value of 0.02 is considered small, 0.15 is considered medium, and 0.35 is considered strong. The results of the effect size test (f^2) showed variations in the strength of influence between constructs in the research model. First, the Anthropomorphism construct has a very small influence on Trust with an f^2 value of 0.008, and contributes almost nothing to User Satisfaction with a value of 0.000. Second, Conversation Capability has a small effect on Trust (0.014) and a moderate effect on User Satisfaction (0.061). Third, the Intention of Continuous Use does not show a significant effect on other constructs, as indicated by its very low f^2 value. Fourth, Information Quality contributes very little to Trust (0.003) but moderately to User Satisfaction (0.053). Fifth, Service Quality has a very small impact on both Trust (0.002) and User Satisfaction (0.001). Sixth, System Quality shows a small impact on Trust (0.044) and a

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 10: 1416-1436, 2025
DOI: 10.55214/2576-8484.v9i10.10674

moderate impact on User Satisfaction (0.136). Next, the Trust path with Intention of Continuous Use has a significant contribution with an f^2 value of 0.263, indicating that Trust plays an important role in explaining the Intention of Continuous Use. Finally, the most dominant path is Trust with User Satisfaction with an f^2 value of 0.951, which indicates a very large and substantial influence. These results show that although most constructs have a small to moderate impact, Trust proved to be the key construct that contributed the most to user satisfaction.

Third, SRMR (Standardized Root Mean Square Residual) to assess model goodness of fit (<0.08) [67].

Table 5.
Test of the Model Fit-SRMR.

	Saturated model	Estimated model
SRMR	0.053	0.054
d_ULS	1.397	1.456
d_G	1.355	1.368
Chi-square	1517.251	1529,528
NFI	0.798	0.796

Table 5 shows a value of 0.053 for the saturated model and 0.054 for the estimated model. This value is smaller than the threshold of 0.08 recommended by Beribisky and Cribbie [67], so it can be concluded that the model has a good fit.. The structural model built in this study is appropriate and suitable for further hypothesis testing.

Fourth, hypothesis testing using Bootstrapping to test the significance of the paths between constructs.

The results of the path coefficient test using the bootstrapping method indicate that not all paths between constructs are statistically significant. The non-significant results include the paths Anthropomorphism to Trust ($p = 0.203$) and Anthropomorphism to User Satisfaction ($p = 0.769$), as well as Information Quality to Trust ($p = 0.608$), Service Quality to Trust ($p = 0.515$), Service Quality to User Satisfaction ($p = 0.867$), and System Quality to Trust ($p = 0.071$). Therefore, it can be concluded that not all dimensions of AI chatbot service quality directly influence user satisfaction and trust. Conversely, several paths demonstrate significant influence, including Conversational Capability to User Satisfaction ($p = 0.007$), Conversational Capability to Trust ($p = 0.025$), Information Quality to User Satisfaction ($p = 0.005$), System Quality to User Satisfaction ($p = 0.006$), and User Satisfaction to Trust ($p = 0.003$).

Additionally, the paths Trust, Intention of Continuous Use ($p = 0.000$), and User Satisfaction, Intention of Continuous Use ($p = 0.000$) showed very strong significance. These findings confirm that Conversational Capability, Information Quality, and System Quality play an important role in increasing user satisfaction and trust, which ultimately drives the intention to continue using chatbot services. These findings confirm that the main factors influencing user satisfaction and trust are Conversational Capability, Information Quality, and System Quality. Ultimately, this satisfaction and trust become the main drivers of users' intention to continue using AI chatbot services. In other words, the natural conversational experience, the quality of relevant information, and the reliability of the system prove to be key determinants in building the sustainability of AI chatbot usage in the future.

Further interpretation suggests that the non-significance of service quality on satisfaction and trust may be due to a shift in user expectations. In the context of AI chatbots, traditional service aspects like quick responses or polite language are already considered the minimum standard, so they are no longer a key differentiating factor. Similarly, anthropomorphism did not show a significant effect because users were aware that chatbots are not human but programmed systems. In fact, under certain conditions, such as when users are frustrated, anthropomorphism can actually decrease satisfaction [68]. Meanwhile, the non-significance of Information Quality and System Quality on Trust indicates that trust is more influenced by emotional experiences than by mere technical aspects.

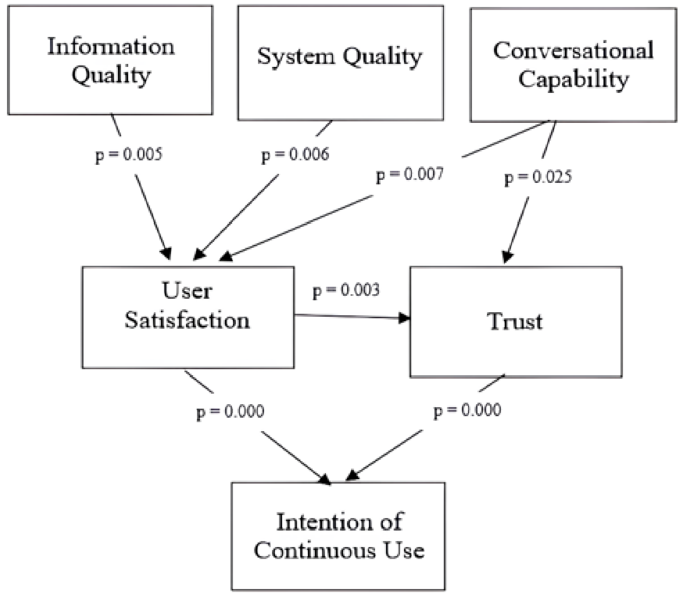

Figure 3.
Optimized Model.

Figure 3 presents an optimized model indicating that not all mediation paths are statistically significant. Significant paths include User Satisfaction, Trust, and Intention of Continuous Use (p = 0.010); System Quality, User Satisfaction, and Trust (p = 0.049); Conversational Capability, User Satisfaction, and Intention of Continuous Use (p = 0.009); Conversational Capability, Trust, and Intention of Continuous Use (p = 0.043); Information Quality, User Satisfaction, and Intention of Continuous Use (p = 0.006); and System Quality, User Satisfaction, and Intention of Continuous Use (p = 0.008). These findings confirm that User Satisfaction and Trust serve as important mediators capable of explaining how some exogenous constructs influence continued use intention. Conversely, insignificant paths include Service Quality, User Satisfaction, and Trust (p = 0.876), and Anthropomorphism, User Satisfaction, and Trust (p = 0.782). This suggests that, in the context of interacting with AI chatbots, traditional service quality aspects and anthropomorphism are not sufficiently strong to build trust through user satisfaction. Therefore, it can be concluded that the continued use of AI chatbots is more mediated by emotional experiences such as user satisfaction and trust, which are formed from the quality of conversations, system reliability, and the relevance of the information provided.

4.7. Norm Development

H1, H6 = Service Quality does not affect User Satisfaction or Trust

The results of this study indicate that service quality does not significantly affect user satisfaction or trust. This finding can be explained by the different nature of AI chatbot services compared to traditional services. Generally, service quality dimensions (such as responsiveness, empathy, and assurance) are more relevant in interactions that involve direct human contact. However, when services are provided by chatbots, users are less likely to evaluate the quality of the interaction based on conventional service aspects and instead place more emphasis on technical factors such as information accuracy, system reliability, and ease of use. In other words, in the context of AI chatbots, system quality and information quality are more dominant in shaping satisfaction and trust than service quality. Additionally, chatbots are seen as self-service tools for obtaining information quickly and accurately, so service quality dimensions like empathy or friendliness are not determining factors in the user

DOI: 10.55214/2576-8484.v9i10.10674

experience [69]. User satisfaction is more determined by the quality of information and reliable systems [34], while trust is formed from the reliability, consistency, and conversational ability of chatbots to provide credible answers [70]. Therefore, these results confirm that in the context of chatbot-based customer service, service quality has a limited influence compared to technical and informational dimensions. In the context of the Telkomsel Veronika Chatbot, it is used for practical needs (checking credit, quota, promotions, and technical services). Users do not pay much attention to service friendliness but are more focused on speed and accuracy of function. Therefore, the service quality dimension is not a determining factor for satisfaction or trust.

H2, H7 = Information Quality influences User Satisfaction, not Trust

Information quality is not significant for trust (p = 0.608), but it is significant for user satisfaction (p = 0.005). This indicates that the quality of information provided by the chatbot is more important for satisfaction than for building trust. This is consistent with the DeLone and McLean [33] where information quality is more closely related to user satisfaction. Customer service chatbot users primarily assess satisfaction based on the relevance, accuracy, and completeness of the information provided. If the information obtained meets needs, satisfaction increases, although this doesn't necessarily build long-term trust. DeLone and McLean [33] emphasize that information quality is closely related to user satisfaction, as good information provides immediate added value. This is also in line with Adam et al. [71] who found that although good information increases satisfaction, user trust is more influenced by the chatbot's ability to maintain coherent conversations and system reliability. Thus, information quality acts as a driver of satisfaction, while trust is more determined by the overall interaction dimension and reliability of the chatbot.

H3, H8 = System Quality significantly influences User Satisfaction but not Trust

System quality significantly influences user satisfaction (p = 0.006), although not trust (p = 0.071). This indicates that the technical reliability, stability, and performance of the chatbot have a more direct impact on user satisfaction than on trust-building. This finding is consistent with the HOT-fit model [18], which places system quality as a determinant of satisfaction. System quality (e.g., response speed, system stability, ease of navigation) has a direct impact on user convenience. If the system is fast and easy to use, users will be satisfied. According to DeLone and McLean [33], system quality is closely related to user satisfaction, as a good technical experience adds value to interactions with information systems. So, even though the system is running well, that's not enough to build trust if users aren't yet convinced of the company's long-term commitment or the accuracy of the service. According to Gefen et al. [72], trust in information systems is more influenced by the integrity of the service provider and the quality of the relationship, rather than solely by the technical aspects of the system. Chatbots are seen as self-service tools. If the system works smoothly, users are satisfied because they feel the process is faster and easier. However, trust is more related to the content of the answers (information quality) and conversational ability, which indicates the chatbot's "good intentions" and competence as a representation of the company. Adam et al. [71] found that system quality increased interaction satisfaction, but trust was more influenced by how well the chatbot maintained a coherent and credible conversation.

H4, H9 = Anthropomorphism did not have a significant impact on either User Satisfaction or Trust.

The analysis results showed that anthropomorphism did not significantly affect trust (p = 0.203) or user satisfaction (p = 0.769), meaning that even though the chatbot has human-like qualities, this did not directly increase user trust and satisfaction. This finding aligns with research [73] stating that anthropomorphism does not necessarily improve users' evaluation of chatbots, especially if the service's primary functions are not met. Additionally, users place more emphasis on functional aspects such as answer quality and service speed compared to the human-like impression of chatbots [74]. In fact, excessive anthropomorphism can cause discomfort, so it does not always contribute to increased trust

[73]. Chatbot AI is essentially a programmed robot. Previous studies have also found that in certain contexts, anthropomorphism is not necessarily a good thing. For example, when customers are angry, anthropomorphism can decrease their satisfaction [68]. In the context of customer service, chatbots are more often seen as tools for resolving issues, so the quality of information and the system are more determining factors in user satisfaction [71]. Thus, the primary focus in developing customer service chatbots should be directed toward conversational ability, information accuracy, and system reliability rather than simply enhancing anthropomorphism.

H5, H10 = Conversational Capability significantly influences User Satisfaction and Trust.

Conversational capability significantly influences trust (p = 0.025) and user satisfaction (p = 0.007). This indicates that a chatbot's ability to maintain natural and interactive conversations is crucial in building user trust and satisfaction. This finding is supported by Gnewuch et al. [69], who emphasize that the quality of conversation is a key determinant of user perception of chatbots. Another finding indicates that the more natural, relevant, and coherent the conversation flow of a chatbot, the higher the level of trust and satisfaction felt by users. Additionally, Ciechanowski et al. [47] found that natural conversational interactions provide a more satisfying user experience compared to rigid conversations. According to Følstad and Skjuve [75], a natural conversation flow enhances the perceived usefulness of chatbots, thereby positively impacting user satisfaction. Consequently, good conversational ability becomes a key factor in creating a trustworthy and satisfying service experience for chatbot users.

H11-H13 = User Satisfaction and Trust as Determinants of Intention of Continuous Use.

User Satisfaction also significantly influences User Satisfaction (p = 0.000), and Trust (p = 0.003) significantly influences the Intention of Continuous Use (p = 0.000). Additionally, this pattern indicates that user satisfaction not only influences the intention to reuse but also increases trust, which in turn strengthens long-term usage commitment. This aligns with the Expectation Confirmation Theory (ECT) model [76], which asserts that satisfaction is a key predictor of continued intention to use information systems. This finding confirms that trust plays a crucial role in users' interactions with customer service AI chatbots, as a sense of trust can reduce uncertainty and provide confidence that the system is reliable, thereby fostering long-term commitment. Users who trust the chatbot's reliability are more likely to intend to continue using it in the future [76] research also supports this finding by showing that trust not only increases satisfaction but is also a key factor in shaping the intention to reuse AI-based digital services. Thus, trust can be seen as the foundation that strengthens the relationship between users and chatbots, both in terms of immediate satisfaction and the sustainability of service use.

5. Discussion

With the rapid development of artificial intelligence (AI) technology, chatbots are no longer just automated answering machines but have evolved into virtual assistants capable of providing interactive and personalized customer service. The sophistication of AI, especially in natural language processing (NLP) and machine learning, makes chatbots increasingly human-like in responding to questions, understanding context, and even adapting their communication style to the user. However, the more advanced the AI technology used, the more urgent the need for a comprehensive evaluation. The findings of this study confirm that the quality of AI chatbots in customer service is primarily determined by technical factors, specifically conversational capability, information quality, and system quality, which have been shown to significantly contribute to user satisfaction and trust. Conversely, the dimensions of anthropomorphism and service quality did not show a significant impact, indicating that users prioritize system reliability and information relevance over human-likeness or service friendliness. From a novelty perspective, this research expands the discourse on AI chatbot evaluation by asserting that the success of chatbots in the context of customer service is not solely determined by anthropomorphic dimensions, as widely assumed in previous studies [73] but rather depends more on conversation quality and system reliability. Additionally, this research integrates various evaluation frameworks—SERVQUAL, the DeLone & McLean IS Success Model, HOT-fit, and AICSQ—into a comprehensive model that is empirically tested in the context of customer service chatbots in the telecommunications industry. Thus, the main contribution of this study lies in emphasizing the importance of functional and interactional dimensions in enhancing user experience, while also offering an evaluation model that is more adaptable to the development of generative AI technology and its application in digital services.

5.1. Theoretical Implications

This research contributes to the study of service quality. First, it bridges the gap between the services provided by AI chatbots and traditional self-service technology (SST) services. While SST generally offers only transactional and functional features, AI chatbots provide a more dynamic and natural conversational form of interaction. Thus, this research expands the theoretical perspective by asserting that user experience with chatbots cannot be fully explained through conventional SST models but rather requires new constructs that emphasize conversational aspects and user trust.

Second, this research integrates dimensions sourced from previous literature, such as service quality, system quality, and information quality (from SERVQUAL, DeLone & McLean, and HOT-fit), while also introducing new constructs like conversational capability. This synthetic approach contributes to the literature by offering a more comprehensive and contextual model for evaluating the quality of AI chatbots. The inclusion of a new dimension confirms that chatbots are not only viewed as information systems but also as conversational entities whose interaction quality directly impacts satisfaction, trust, and intention to continue using them.

Third, this research expands and fills gaps in the AICSQ (Artificial Intelligence Customer Service Quality) model, which previously treated the human-like dimension as a single variable. This study breaks it down into two separate constructs: anthropomorphism and conversational capability. This separation makes an important conceptual contribution because it shows that human-like impressions do not always align with technical conversational ability. Thus, the chatbot quality evaluation model becomes more specific, operational, and relevant to the context of digital customer service.

Fourth, these findings strengthen the position of User Satisfaction and Trust as key mediators in the AI chatbot quality evaluation model. Their roles indicate that the relationship between technical and psychological quality and continued usage intention is not direct but rather mediated by users' subjective experiences. This aligns with the satisfaction and trust theory in information systems, which emphasizes the importance of affective factors in bridging service quality and user behavior [76]. Thus, this research contributes to the literature by confirming that the quality evaluation of AI chatbots must include satisfaction and trust mediation mechanisms as core components in its conceptual framework.

5.2. Organizational Implications

This research has important implications for organizations involved in the AI chatbot ecosystem. First, for AI chatbot developers, the research findings can be used as a benchmark to understand the key dimensions of service quality. This understanding allows developers to assess whether the resources they possess, both in terms of technology, design, and infrastructure, are sufficient to produce a chatbot with high service quality. Thus, this research supports a more directed and standardized chatbot production process that meets user expectations. Second, for organizations using AI chatbots, these findings can serve as a basis for evaluating the service quality level of the chatbots they implement. Through the tested indicators, organizations can assess whether the chatbot used is truly capable of meeting customer needs, increasing satisfaction, building trust, and reusing the AI chatbot. This evaluation can also help organizations formulate improvement strategies and further development, so that the chatbot functions not only as an automated service tool but also as an effective and sustainable digital interaction channel in supporting customer experience. Thirdly, for customers or end-users, the results of this research confirm that with clear evaluation models, customers can be more critical in assessing the chatbots they use, while also having realistic expectations regarding the quality of AI-based services. This ultimately fosters a more balanced relationship between service providers and users, where customers gain a satisfying digital experience, while organizations receive valuable feedback for continuous improvement.

5.3. Limitations and Future Research

This study has several limitations that open opportunities for future research. First, it focuses on one type of AI chatbot in the context of customer service with experiments in the telecommunications industry. However, various types of AI chatbots are used in other sectors, such as e-commerce, banking, healthcare, and education. Each context has different service characteristics, leading to variations in user needs, expectations, and quality criteria. Therefore, future research can expand the conceptualization and measurement of AI chatbot service quality by adapting to a wider range of contexts and chatbot types.

Second, the dimensions of service quality developed in this study are adapted from the SERVQUAL model theory, the DeLone & McLean IS Success Model, HOT-fit, and AICSQ. These four models are sufficiently representative to measure the quality of information, systems, and technology services. Although these models have been widely used and proven relevant, the selection of dimensions in this study remains contextual, based on the characteristics of the AI chatbot being researched. Thus, this study has limitations in terms of generalizing the service quality dimensions to other contexts. Future studies are recommended to explore and test additional dimensions that may be more relevant, either by referring to previous theoretical models such as E-S-QUAL or by developing new dimensions specific to the chatbot's main goals and tasks. This approach is expected to broaden understanding of the quality of AI chatbot services across industries and functions.

Third, this study found that the dimensions of anthropomorphism and service quality did not significantly affect user satisfaction or trust, although previous literature indicated that human-like aspects influence user satisfaction [42, 55] and the importance of service quality for customer loyalty [14]. This finding raises new questions about how the impact and internal mechanisms of human-like chatbots function in various service contexts. Therefore, further research is needed to delve deeper into the role and limitations of anthropomorphic features and to explore whether their effects differ across specific sectors or service types.

Fourth, this research confirms that user satisfaction and trust play a key mediating role in building continued use intention. Nevertheless, the focus of this research is still limited to these two variables and does not yet accommodate other factors that theoretically have the potential to influence continued use. Research in the telecommunications industry at Telecom Egypt shows that trust can improve the overall effectiveness of chatbots [77]. Several relevant factors, such as perceived value (the benefits gained compared to the effort/cost expended in using the chatbot), are worth considering in future research to enrich the conceptual model. Additionally, further research is also suggested to compare the roles of these variables in various industrial contexts. For example, in knowledge-based organizations, perceived value is likely to be a more prominent determining factor, whereas in the context of e-commerce, user satisfaction tends to be more dominant. Thus, integrating additional factors and exploring cross-industry contexts can provide a more comprehensive understanding of the mechanisms underlying the formation of the intention to use sustainable AI chatbots.

Overall, future research needs to develop more contextualized AI chatbot service quality evaluation models, expanding the integration of psychological and technical factors. Considering that modern AI-based chatbots, NLP (Natural Language Processing), and LLMs (Large Language Models) can introduce new characteristics to chatbots, psychological and technical factors need to be integrated. With this direction, future studies can enrich theoretical understanding while also making broader practical contributions to the development and implementation of cross-industry AI chatbots.

6. Conclusion and Recommendation

This research successfully developed and tested a service quality measurement model for AI chatbots in the context of customer service, using the Telkomsel Veronika chatbot as a case study. The analysis results show that not all dimensions of chatbot service quality significantly affect user satisfaction and trust. The variables of Anthropomorphism and Service Quality were proven to have no significant impact on User Satisfaction or Trust, confirming that users prioritize functional and technical aspects over human-like service impressions. Conversely, Conversational Capability, Information Quality, and System Quality were proven to have a significant impact on User Satisfaction, which in turn plays a crucial role in increasing Trust and driving the intention of continuous use. Additionally, Trust also has a direct role in strengthening satisfaction and the intention of continuous use. This aligns with the expectations of further research by Chen et al. [42] to explore how trust factors can influence continued use intention. Theoretically, this research contributes to the development of a chatbot quality evaluation model by emphasizing the integration of technical and conversational aspects as key dimensions. Empirically, this research strengthens the evidence that satisfaction and trust are major determinants in the successful implementation of AI chatbots in customer service for repeat chatbot usage.

This research provides a strong conceptual foundation for the development of further studies on the evaluation of user experience-based chatbot quality. First, future research is suggested to experimentally test the role of Conversational Capability as a mediating variable in the relationship between Anthropomorphism and user experience outcomes, such as satisfaction and trust. This experimental approach will provide a deeper understanding of the influence of "human-like" perception on conversation quality. Second, it is important for future research to consider local language and cultural adaptation factors, as perceptions of human interaction and conversation quality are heavily influenced by the communication norms and social expectations specific to each user context. Third, future research needs to expand the context to various types of chatbots, such as in the education sector, e-commerce, and government public services, to test the model's consistency across different domains. Thus, this research not only contributes theoretically to the development of a chatbot quality evaluation model but also offers relevant practical implications for improving AI-based customer service.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis, "Transforming the communication between citizens and government through AI-guided chatbots," *Government Information Quarterly*, vol. 36, no. 2, pp. 358-367, 2019. https://doi.org/10.1016/j.giq.2018.10.001

[2] A. G. Larsen and A. Følstad, "The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers," *Government Information Quarterly*, vol. 41, no. 2, p. 101927, 2024. https://doi.org/10.1016/j.giq.2024.101927

[3] A. De Keyser and W. H. Kunz, "Living and working with service robots: A TCCM analysis and considerations for future research," *Journal of Service Management*, vol. 33, no. 2, pp. 165-196, 2022. https://doi.org/10.1108/JOSM-12-2021-0488

[4] M. E. Cortés-Cediel, A. Segura-Tinoco, I. Cantador, and M. P. R. Bolívar, "Trends and challenges of e-government chatbots: Advances in exploring open government data and citizen participation content," *Government Information Quarterly*, vol. 40, no. 4, p. 101877, 2023. https://doi.org/10.1016/j.giq.2023.101877

[5] G. Todorov, "117 cool chatbot statistics 2024 [Learn Digital Marketing]. Thrivemyway," 2023. https://thrivemyway.com/chatbot-statistics/. [Accessed Feb. 05, 2024]

[6] P. Vassilakopoulou, A. Haug, L. M. Salvesen, and I. O. Pappas, "Developing human/AI interactions for chat-based customer services: Lessons learned from the Norwegian government," *European Journal of Information Systems*, vol. 32, no. 1, pp. 10-22, 2023. https://doi.org/10.1080/0960085X.2022.2096490

[7] M. B. S. Prasad and B. Reddemma, "Automatic E-governing system using ML & AI," *International Journal of Engineering Technology and Management Sciences*, vol. 6, no. 5, pp. 632–644, 2022. https://doi.org/10.46647/ijetms.2022.v06i05.100
[8] S. Dutta, "Chatbot effectiveness in enhancing guest communication: Insights from secondary data," *Tuijin Jishu / Journal of Propulsion Technology*, vol. 45, no. 4, pp. 2137–2150, 2024. https://doi.org/10.52783/tjjpt.v45.i04.8497
[9] J. J. A. Ocxas, F. K. C. Martin, and A. N. H. Junes, "AI chatbot with emotional recognition for an improved customer service experience," in *2024 IEEE XXXI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, 2024: IEEE, pp. 1-8.
[10] M. Al-Shafei, "Navigating human-chatbot interactions: An investigation into factors influencing user satisfaction and engagement," *International Journal of Human–Computer Interaction*, vol. 41, no. 1, pp. 411-428, 2025. https://doi.org/10.1080/10447318.2023.2301252
[11] J. Zhang, Q. Chen, J. Lu, X. Wang, L. Liu, and Y. Feng, "Emotional expression by artificial intelligence chatbots to improve customer satisfaction: Underlying mechanism and boundary conditions," *Tourism Management*, vol. 100, p. 104835, 2024. https://doi.org/10.1016/j.tourman.2023.104835
[12] Y. Zhu, J. Zhang, and J. Liang, "Concrete or abstract: How chatbot response styles influence customer satisfaction," *Electronic Commerce Research and Applications*, vol. 62, p. 101317, 2023. https://doi.org/10.1016/j.elerap.2023.101317
[13] F. Thomaz, C. Salge, E. Karahanna, and J. Hulland, "Learning from the Dark Web: Leveraging conversational agents in the era of hyper-privacy to enhance marketing," *Journal of the Academy of Marketing Science*, vol. 48, pp. 43-63, 2020. https://doi.org/10.1007/s11747-019-00704-3
[14] Q. Chen, Y. Lu, Y. Gong, and J. Xiong, "Can AI chatbots help retain customers? Impact of AI service quality on customer loyalty," *Internet Research*, vol. 33, no. 6, pp. 2205-2243, 2023. https://doi.org/10.1108/INTR-09-2021-0686
[15] M. Song, H. Zhang, X. Xing, and Y. Duan, "Appreciation vs. apology: Research on the influence mechanism of chatbot service recovery based on politeness theory," *Journal of Retailing and Consumer Services*, vol. 73, p. 103323, 2023. https://doi.org/10.1016/j.jretconser.2023.103323
[16] A. Parasuraman, V. A. Zeithaml, and L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, vol. 64, no. 1, pp. 12-40, 1988.
[17] A. Zuiderwijk, Y.-C. Chen, and F. Salem, "Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda," *Government Information Quarterly*, vol. 38, no. 3, p. 101577, 2021. https://doi.org/10.1016/j.giq.2021.101577
[18] M. M. Yusof, J. Kuljis, A. Papazafeiropoulou, and L. K. Stergioulas, "An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit)," *International Journal of Medical Informatics*, vol. 77, no. 6, pp. 386-398, 2008. https://doi.org/10.1016/j.ijmedinf.2007.08.011
[19] E. E. Sala and A. P. Subriadi, "Hot-Fit model to measure the effectiveness and efficiency of information system in public sector," *The Winners*, vol. 23, no. 2, pp. 131–141, 2023. https://doi.org/10.21512/tw.v23i2.7423
[20] L. M. Erlirianto, A. H. N. Ali, and A. Herdiyanti, "The implementation of the Human, Organization, and Technology–Fit (HOT–Fit) framework to evaluate the electronic medical record (EMR) System in a Hospital," *Procedia Computer Science*, vol. 72, pp. 580-587, 2015. https://doi.org/10.1016/j.procs.2015.12.166
[21] B. Magnusson, "Measurement quality in analysis—guidelines and software tools," in *International Conference EcoBalt 2023 Chemicals & Environment*, 2023.
[22] L. Liu, S. Isaacman, and U. Kremer, "An adaptive application framework with customizable quality metrics," *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, vol. 27, no. 2, pp. 1-33, 2021. https://doi.org/10.1145/3477428
[23] N. Aoki, "An experimental study of public trust in AI chatbots in the public sector," *Government information quarterly*, vol. 37, no. 4, p. 101490, 2020. https://doi.org/10.1016/j.giq.2020.101490
[24] A. S. Lubis, Y. Yulinda, A. Qaedi Hutagalung, A. Dilham, F. R. Sofiyah, and J. L. Marpaung, "The chatbot artificial intelligence as the alternative customer services strategic to improve the customer relationship management in real-time responses," *International Journal of Economics and Business Research*, vol. 27, no. 5, pp. 45–58, 2024.
[25] S. Barnes and R. Vidgen, "WebQual: An exploration of website quality," in *Proceedings of the 8th European Conference on Information Systems (ECIS 2000) (pp. 298–305). Vienna, Austria*, 2000.
[26] M. Aljukhadar, J.-F. Belisle, D. C. Dantas, S. Sénécal, and R. Titah, "Measuring the service quality of governmental sites: Development and validation of the e-Government service quality (EGSQUAL) scale," *Electronic Commerce Research and Applications*, vol. 55, p. 101182, 2022. https://doi.org/10.1016/j.elerap.2022.101182
[27] A. Jonkisz, P. Karniej, and D. Krasowska, "SERVQUAL method as an "old new" tool for improving the quality of medical services: A literature review," *International Journal of Environmental Research and Public Health*, vol. 18, no. 20, p. 10758, 2021. https://doi.org/10.3390/ijerph182010758
[28] F. L. Lizarelli, L. Osiro, G. M. Ganga, G. H. Mendes, and G. R. Paz, "Integration of SERVQUAL, Analytical Kano, and QFD using fuzzy approaches to support improvement decisions in an entrepreneurial education service," *Applied Soft Computing*, vol. 112, p. 107786, 2021. https://doi.org/10.1016/j.asoc.2021.107786

[29] E. Tumsekcali, E. Ayyildiz, and A. Taskin, "Interval valued intuitionistic fuzzy AHP-WASPAS based public transportation service quality evaluation by a new extension of SERVQUAL Model: P-SERVQUAL 4.0," *Expert Systems with Applications*, vol. 186, p. 115757, 2021. https://doi.org/10.1016/j.eswa.2021.115757
[30] P. Bhattacharya *et al.*, "Perception-satisfaction based quality assessment of tourism and hospitality services in the Himalayan region: An application of AHP-SERVQUAL approach on Sandakphu Trail, West Bengal, India," *International Journal of Geoheritage and Parks*, vol. 11, no. 2, pp. 259-275, 2023. https://doi.org/10.1016/j.ijgeop.2023.04.001
[31] L. Ocampo *et al.*, "Public service quality evaluation with SERVQUAL and AHP-TOPSIS: A case of Philippine government agencies," *Socio-Economic Planning Sciences*, vol. 68, p. 100604, 2019. https://doi.org/10.1016/j.seps.2017.12.002
[32] S. Afroj *et al.*, "Assessing the municipal service quality of residential neighborhoods based on SERVQUAL, AHP and Citizen's Score Card: A case study of Dhaka North City Corporation area, Bangladesh," *Journal of Urban Management*, vol. 10, no. 3, pp. 179-191, 2021. https://doi.org/10.1016/j.jum.2021.03.001
[33] W. H. DeLone and E. R. McLean, "The DeLone and McLean model of information systems success: A ten-year update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9-30, 2003. https://doi.org/10.1080/07421222.2003.11045748
[34] M. Shim and H. S. Jo, "What quality factors matter in enhancing the perceived benefits of online health information sites? Application of the updated DeLone and McLean Information Systems Success Model," *International Journal of Medical Informatics*, vol. 137, p. 104093, 2020. https://doi.org/10.1016/j.ijmedinf.2020.104093
[35] M. S. Hassan, M. A. Islam, M. F. Yusof, and H. Nasir, "Users' fintech services acceptance: A cross-sectional study on Malaysian Insurance & takaful industry," *Heliyon*, vol. 9, no. 11, p. e21130, 2023. https://doi.org/10.1016/j.heliyon.2023.e21130
[36] A. Lutfi, "Factors affecting the success of accounting information system from the lens of DeLone and McLean IS model," *International Journal of Information Management Data Insights*, vol. 3, no. 2, p. 100202, 2023. https://doi.org/10.1016/j.jjimei.2023.100202
[37] C. Tam and T. Oliveira, "Understanding the impact of m-banking on individual performance: DeLone & McLean and TTF perspective," *Computers in Human Behavior*, vol. 61, pp. 233-244, 2016. https://doi.org/10.1016/j.chb.2016.03.016
[38] A. H. Aldholay, O. Isaac, Z. Abdullah, and T. Ramayah, "The role of transformational leadership as a mediating variable in DeLone and McLean information system success model: The context of online learning usage in Yemen," *Telematics and Informatics*, vol. 35, no. 5, pp. 1421-1437, 2018. https://doi.org/10.1016/j.tele.2018.03.012
[39] S. Zheng, Z. Yahya, L. Wang, R. Zhang, and A. N. Hoshyar, "Multiheaded deep learning chatbot for increasing production and marketing," *Information Processing & Management*, vol. 60, no. 5, p. 103446, 2023. https://doi.org/10.1016/j.ipm.2023.103446
[40] W. Liu, M. Jiang, W. Li, and J. Mou, "How does the anthropomorphism of AI chatbots facilitate users' reuse intention in online health consultation services? The moderating role of disease severity," *Technological Forecasting and Social Change*, vol. 203, p. 123407, 2024. https://doi.org/10.1016/j.techfore.2024.123407
[41] Y. Xu, N. Niu, and Z. Zhao, "Dissecting the mixed effects of human-customer service chatbot interaction on customer satisfaction: An explanation from temporal and conversational cues," *Journal of Retailing and Consumer Services*, vol. 74, p. 103417, 2023. https://doi.org/10.1016/j.jretconser.2023.103417
[42] Q. Chen, Y. Gong, Y. Lu, and J. Tang, "Classifying and measuring the service quality of AI chatbot in frontline service," *Journal of Business Research*, vol. 145, pp. 552-568, 2022. https://doi.org/10.1016/j.jbusres.2022.02.088
[43] G. A. Churchill Jr, "A paradigm for developing better measures of marketing constructs," *Journal of Marketing Research*, vol. 16, no. 1, pp. 64-73, 1979. https://doi.org/10.2307/3150876
[44] J. C. Anderson and D. W. Gerbing, "Structural equation modeling in practice: A review and recommended two-step approach," *Psychological Bulletin*, vol. 103, no. 3, pp. 411-423, 1988. https://doi.org/10.1037/0033-2909.103.3.411
[45] J. F. Hair, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM)*, 2nd ed. Thousand Oaks, CA: SAGE Publications, 2017.
[46] N. Cai, S. Gao, and J. Yan, "How the communication style of chatbots influences consumers' satisfaction, trust, and engagement in the context of service failure," *Humanities and Social Sciences Communications*, vol. 11, p. 687, 2024. https://doi.org/10.1057/s41599-024-03212-0
[47] L. Ciechanowski, A. Przegalinska, M. Magnuski, and P. Gloor, "In the shades of the uncanny valley: An experimental study of human–chatbot interaction," *Future Generation Computer Systems*, vol. 92, pp. 539-548, 2019. https://doi.org/10.1016/j.future.2018.01.055
[48] M. Hmoud, H. Swaity, E. Anjass, and E. M. Aguaded-Ramírez, "Rubric development and validation for assessing tasks' solving via AI chatbots," *Electronic Journal of e-Learning*, vol. 22, no. 6, pp. 1-17, 2024. https://doi.org/10.34190/ejel.22.6.3292
[49] M. J. A. Craig, C. Edwards, and A. Edwards, *Human-computer interaction*. Berlin/Boston: De Gruyter, 2025.
[50] H. T. Reyburn, "Human–computer interaction," *International Encyclopedia of Health Communication*, pp. 1–7, 2022.

[51] G. R. S. Silva and E. D. Canedo, "Unveiling quality in chatbot conversations: Quantitative analysis of chatbot requirements," in *Proceedings of the XXII Brazilian Symposium on Software Quality*, 2023.

[52] R. Khan, S. P. Khan, and S. A. Ali, "Conversational AI: Dialoguing most humanly with non-humans," *Conversational Artificial Intelligence*, pp. 249-268, 2024. https://doi.org/10.1002/9781394200801.ch16

[53] C. G. Møller, K. E. Ang, M. de Lourdes Bongiovanni, M. S. Khalid, and J. Wu, "Metrics of success: Evaluating user satisfaction in AI chatbots," in *Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence*, 2024, pp. 168-173.

[54] M. Ashfaq, J. Yun, S. Yu, and S. M. C. Loureiro, "I, Chatbot: Modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents," *Telematics and Informatics*, vol. 54, p. 101473, 2020. https://doi.org/10.1016/j.tele.2020.101473

[55] N. Noor, S. Rao Hill, and I. Troshani, "Developing a service quality scale for artificial intelligence service agents," *European Journal of Marketing*, vol. 56, no. 5, pp. 1301-1336, 2022. https://doi.org/10.1108/EJM-09-2020-0672

[56] J. Zhang, J. Lu, X. Wang, L. Liu, and Y. Feng, "Emotional expressions of care and concern by customer service chatbots: Improved customer attitudes despite perceived inauthenticity," *Decision Support Systems*, vol. 186, p. 114314, 2024. https://doi.org/10.1016/j.dss.2024.114314

[57] C.-L. Hsu and J. C.-C. Lin, "Understanding the user satisfaction and loyalty of customer service chatbots," *Journal of Retailing and Consumer Services*, vol. 71, p. 103211, 2023. https://doi.org/10.1016/j.jretconser.2022.103211

[58] X. Li, S. Xie, Z. Ye, S. Ma, and G. Yu, "Investigating patients' continuance intention toward conversational agents in outpatient departments: Cross-sectional field survey," *Journal of Medical Internet Research*, vol. 24, no. 11, p. e40681, 2022. https://doi.org/10.2196/40681

[59] J. Kim and K. Yum, "Enhancing continuous usage intention in e-commerce marketplace platforms: The effects of service quality, customer satisfaction, and trust," *Applied Sciences*, vol. 14, no. 17, p. 7617, 2024. https://doi.org/10.3390/app14177617

[60] J. Li and Y. F. Lay, "Examining the reliability and validity of self-efficacy beliefs, stress, perceived teachers' support and academic burnout scales using the PLS-SEM approach," *Humanities and Social Sciences Letters*, vol. 13, no. 1, pp. 56-68, 2025. https://doi.org/10.18488/73.v13i1.3987

[61] N. K. Avkiran, "An in-depth discussion and illustration of partial least squares structural equation modeling in health care," *Health Care Management Science*, vol. 21, no. 3, pp. 401-408, 2018. https://doi.org/10.1007/s10729-017-9393-7

[62] K. D. Carlson and A. O. Herdman, "Understanding the impact of convergent validity on research results," *Organizational Research Methods*, vol. 15, no. 1, pp. 17-32, 2012. https://doi.org/10.1177/1094428110392383

[63] J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *Journal of the Academy of Marketing Science*, vol. 43, no. 1, pp. 115-135, 2015. https://doi.org/10.1007/s11747-014-0403-8

[64] E. Roemer, F. Schuberth, and J. Henseler, "HTMT2–an improved criterion for assessing discriminant validity in structural equation modeling," *Industrial Management & Data Systems*, vol. 121, no. 12, pp. 2637-2650, 2021. https://doi.org/10.1108/IMDS-02-2021-0082

[65] J. F. Hair Jr, G. T. M. Hult, C. M. Ringle, M. Sarstedt, N. P. Danks, and S. Ray, *Evaluation of the structural model*. Cham: Springer, 2021.

[66] J. Cohen, "Statistical power analysis," *Current Directions in Psychological Science*, vol. 1, no. 3, pp. 98-101, 1992.

[67] N. Beribisky and R. A. Cribbie, "Equivalence testing based fit index: Standardized root mean squared residual," *Multivariate Behavioral Research*, vol. 60, no. 1, pp. 138-157, 2025. https://doi.org/10.1080/00273171.2024.2386686

[68] C. Crolic, F. Thomaz, R. Hadi, and A. T. Stephen, "Blame the bot: Anthropomorphism and anger in customer–chatbot interactions," *Journal of Marketing*, vol. 86, no. 1, pp. 132-148, 2022. https://doi.org/10.1177/00222429211045687

[69] U. Gnewuch, S. Morana, and A. Maedche, "Towards designing cooperative and social conversational agents for customer service," presented at the ICIS 2017: Transforming Society with Digital Innovation, 2017.

[70] J. Mou and M. Benyoucef, "Consumer behavior in social commerce: Results from a meta-analysis," *Technological Forecasting and Social Change*, vol. 167, p. 120734, 2021. https://doi.org/10.1016/j.techfore.2021.120734

[71] M. Adam, M. Wessel, and A. Benlian, "AI-based chatbots in customer service and their effects on user compliance," *Electronic Markets*, vol. 31, no. 2, pp. 427-445, 2021. https://doi.org/10.1007/s12525-020-00414-7

[72] D. Gefen, E. Karahanna, and D. W. Straub, "Trust and TAM in online shopping: An integrated model," *MIS Quarterly*, vol. 27, no. 1, pp. 51-90, 2003.

[73] S. Moussawi, M. Koufaris, and R. Benbunan-Fich, "How perceptions of intelligence and anthropomorphism affect adoption of personal intelligent agents," *Electronic Markets*, vol. 31, no. 2, pp. 343-364, 2021. https://doi.org/10.1007/s12525-020-00411-w

[74] E. Go and S. S. Sundar, "Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions," *Computers in Human Behavior*, vol. 97, pp. 304-316, 2019. https://doi.org/10.1016/j.chb.2019.01.020

[75] A. Følstad and M. Skjuve, "Chatbots for customer service: User experience and motivation," in *Proceedings of the 1st International Conference on Conversational User Interfaces*, 2019, pp. 1-9.

[76] A. Bhattacherjee, "Understanding information systems continuance: An expectation-confirmation model," *MIS Quarterly*, vol. 25, no. 3, pp. 351-370, 2001.

[77] A. Tigariyyat and M. A. D. Al Maliyyat, "Impact of AI chatbots on customer engagement: An investigating in the role of customer trust in online services of telecom egypt context," *Journal of Capital Markets Studies*, vol. 34, no. 3, pp. 1–42, 2024.

ISSN: 2576-8484
Vol. 9, No. 10: 1416-1436, 2025
DOI: 10.55214/2576-8484.v9i10.10674