

Review Article

Trustworthy AI in Customer Service: A Conceptual Framework for Building Customer Trust

Patrick Monis^{1*}, Prof. (Dr.) Peeyush Kumar Pandey²¹Research Scholar (Ph.D. Candidate), Department of Psychology, Faculty of Humanities, Languages & Social Sciences, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India²Professor, School of Commerce and Management, Shri Venkateshwara University, Gajraula, Uttar Pradesh, India**Article History**

Received: 24.02.2026

Accepted: 27.03.2026

Published: 29.04.2026

Journal homepage:<https://www.easpublisher.com>**Quick Response Code**

Abstract: AI has rapidly become a key player in the Customer Service landscape, revolutionizing how companies interact with customers, deliver support, personalize experiences, and navigate customer interactions. The use of AI-driven chatbots, virtual assistants, recommendation systems, predictive analytics, and generative AI are driving the shift of traditional service provider roles into automated tasks. While the technologies offer great benefits in terms of responsiveness, availability, personalization, and operational efficiency, achieving the adoption of these technologies greatly relies on customer trust. Fears and concerns about privacy, security, fairness, wrong answers, lack of transparency, or no human involvement may prevent customers from engaging with AI. Thus, a trusted AI-driven customer service goes beyond just technical capabilities. This research identifies and forays into an empirical model to explore the effect of the trustworthy attributes of AI on consumer trust in AI-driven customer service. Transparency and explainability, fairness, privacy and data protection, reliability and robustness, security, human oversight, and accountability are the main elements of the proposed framework for trustworthy AI. These dimensions are recommended to be viewed as a higher-order construct that is perceived trustworthiness of AI that impacts customer trust. Perceived risk is included as an intervening variable and perceived customer control is a trust-enhancing complimentary factor. Customer trust is then suggested to affect customer satisfaction, continuance intention, and customer loyalty. The research design is quantitative research which is applied to the respondents who have previously contacted the customer-service systems based on artificial intelligence. The study is of quantitative research method with a structured questionnaire to the customers who have previously interacted with the customer-service systems based on artificial intelligence. Measurement and structural models are recommended to be explored through structural equations modelling and direct, indirect and moderating relationships are recommended to be tested. The proposed study extends the trustworthy-AI and customer-service literature to the customer-level psychological or behavioral outcomes by examining how the characteristics of AI governance relate to these outcomes. The framework can help organizations to create technically reliable, ethically responsible, transparent, secure, and human-centered AI systems for customer service.

Keywords: Trustworthy AI, Artificial Intelligence, Customer Service, Customer Trust, Explainable AI, Transparency, Privacy, Fairness, Human Oversight, Perceived Risk, Customer Satisfaction, Customer Loyalty.

Copyright © 2026 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)** which permits unrestricted use, distribution, and reproduction in any medium for non-commercial use provided the original author and source are credited.

INTRODUCTION

Artificial intelligence is becoming one of the most significant technologies that impact today's customer-service management [1]. AI systems are becoming more prevalent in the financial sector and are used by various organizations in the banking, telecommunications, retail, healthcare, travel, e-commerce, insurance, and other service sectors to interact with customers and streamline service operations. The AI-powered customer-service technologies are able to respond to customer demands, categorize service requests, understand customer intent, suggest products, gauge a customer's sentiment, find pertinent information, and compose a personalized reply [2]. This has been further enhanced by the growing presence of Generative AI, which enables the involvement of AI in engaging in natural-language conversations and increasingly elaborated service tasks. AI's use in customer service offers a number of benefits to the operations of a business. AI systems have the ability to be open around-the-clock, react quickly to many customers, cut down repetitive work for employees, and even enhance personalization. AI can help the customer without the need to wait for a human to serve them; it can give them immediate access to information and service [3]. But as AI systems become more independent and complex, there are also many trust issues. Customer service often involves the divulging of personal and commercial sensitive information. During AI conversations, customers can submit their name, contact details, transaction data, preferences, grievances, payment data, and other data related to their behavior. Therefore, customers need to trust the AI system that will safeguard their data, as well as trust the organization in how they use that data [4]. Moreover, customers can also have doubts about the accuracy of AI-generated responses, fairness of automated decisions, ability to understand the decision-making process, and the possibility of human support when AI is not successful. These worries highlight that customer acceptance is not solely about the performance of AI. Customers' mistrust of the system, even when it is fast, can occur if the system is intrusive or unreliable. Likewise, a system can be technologically advanced but be rejected if the customer feels that decisions made via automation are unfair and/or not subject to appeal. As such, trustworthy AI is a key pillar for the sustainable use of AI in customer service [5], [6].

Reliable AI has many facets – it is reliable, safe, secure, transparent, explainable, private, fair, and accountable. The dimensions have been increasingly included in international frameworks of AI governance. But there is a big research gap in the connection between trustworthy-AI principles and customer outcomes [7], [8]. The technical and regulatory conversations tend to focus on the nature of trustworthy AI, and research into customer service tends to be more about the quality of service, its usefulness, convenience and satisfaction. Compared to this, the integration of these aspects into a

consolidated empirical model that looks into how these characteristics of AI contribute to customer trust is comparatively less [9], [10].

Research Gaps

The central research gap is that there is a lack of an empirical model with enough differentiation in terms of how customer-experienced trustworthy-AI characteristics are translated into customer trust via perceived AI trustworthiness and perceived risk, with perceived customer control as a boundary condition, and involving downstream customer outcomes. To fill this gap, this study introduces a trustworthy-AI framework with a customer-centric perspective. The study will specifically explore the effect of transparency/explainability, fairness, privacy and data protection, reliability/robustness, security, human oversight, and accountability on their perception of the trustworthiness of AI and customer trust. It also reviews the possibility of perceived risk moderating the relationship between trustworthy AI and customer trust, and whether customer trust relates to customer satisfaction, continuance intention, and customer loyalty. This study aims to create and test an extensive framework that elucidates the mechanisms of customer trust formation in AI-based customer service.

The study aims at answering the following research questions:

- **RQ1:** What are the effects of the characteristics of trustworthy-AI on the perceived trustworthiness in customer service?
- **RQ2:** How does customer trust relate to perceived trustworthiness of AI?
- **RQ3:** Does the relationship between trustworthy AI and customer trust go through perceived risk?
- **RQ4:** Can a sense of customer control enhance customer confidence in AI-powered customer service?
- **RQ5:** What is the relationship between customer trust and customer satisfaction, continuance intention and customer loyalty?

RELATED WORK

Trustworthy, Ethical, and Governed AI Studies

Concepts of openness, responsibility, equity, dependability, and confidentiality are central to the work of Trustworthy AI in Tourism that was outlined in that article. The article presents the majority of the examples and ethical issues associated with the adoption of AI in the tourist industry by evaluating its applications in areas such as customer service, predictive analytics, personalized recommendations, while familiarization with the location. The study delves into real-life examples and recommended practices to ensure the AI gadgets are dependable. Some of the key ways include establishing and adhering to ethical principles, establishing channels for transparency, and communicating with stakeholders. That research

presents conclusions based on a comprehensive analysis of the present and potential future states of affairs; these states demonstrate the integration of human-trustworthy AI and encourage the growth of the tourist sector [11]. The research delves into the ethical adoption of AI in the service sectors via a systematic literature review paired with bibliometric analysis. Important issues including privacy, prejudice, transparency, and accountability are brought to light in that study, highlighting the urgent need for ethical AI methods in service industries that deal with sensitive consumer data. In domains including as decision assistance, customer interaction, automation, along with new service creation, the results show that the ethical use of AI is critical. Ethical standards, human supervision, thorough training, and a flexible organizational culture are all enablers that the study reveals are crucial for releasing AI's potential while cutting hazards. In order to create trustworthy and responsible AI for the service sectors, the study lays forth a plan for future research that includes multidisciplinary cooperation, consumer co-creation in moral structures, and sector-specific policy adaption [18].

Data bias, privacy concerns, and governance are some of the ethical aspects of artificial intelligence that that article focuses on. That study presents a hybrid governance approach that draws on many main frameworks to tackle the specific issues of operational settings. These frameworks include the OECD AI Principles along with the EUs Ethics Guidelines for Trustworthy AI. The utilization of a financial industry case study deepens the understanding of how privacy-preserving strategies may protect AI-driven customer service's sensitive client data. Extensive testing in that instance shown that privacy-preserving techniques, such as federated learning and differential privacy, may increase customer happiness by over 20% and decrease the occurrence of unauthorised data-access incidents by 30%. By providing businesses with actionable advice on how to responsibly and legally operate AI operations, that article adds to the lively conversation around ethical AI [19]. Having said that, end-user awareness is still poor. The government and businesses are still exploring potential solutions to the problems that have been brought to light in order to regulate, optimize, and harmonize them. To create reliable AI-powered systems, developers and businesses must do more than just adhere to broad regulations; they must also think about ethical concerns. The goal of that article is not to provide broad answers but rather to construct a conceptual framework through studying potential solutions within existing debates, concerns, theories, policies, and difficulties, as well as by incorporating the author's own personal perspective. The organization's opinions may differ from those expressed in that document [20].

AI-Driven Customer Experience, Chatbots, and Personalization

In order to create reliable Zero-Touch Customer Experience systems, that research suggests a

Governance-Constrained Agentic AI Framework. Autonomous process optimization techniques, explainable AI applications, real-time observability pipelines, adaptive governance layers, policy-aware decision engines, and multi-agent orchestration are all part of the system. Full-service autonomy is supported by the architecture, which also ensures accountability, transparency, regulatory compliance, and data governance. Intelligent autonomous services, strict enforcement of governance, adaptive orchestration, and constant monitoring of compliance make up the suggested architecture. The following features are included: risk scoring, audit trails, semantic validation, identifying anomalies, contextual reasoning, behavioral analytics, adaptive personalization, regulations regarding ethical AI, and consumer intent prediction. In order to optimize client interactions on the fly, distributed AI agents work under governance limitations by using reinforcement learning models and hierarchical orchestration. Compared to conventional AI-based customer service systems, the experimental assessment shows considerable benefits, such as a 38% increase in operational efficiency, a 41% decrease in service escalations, a 32% increase in customer satisfaction, and a 57% decrease in compliance violations. Enterprise resilience in multi-cloud contexts, service continuity, adaptive transparency, with observability accuracy are all improved by the architecture. In order to provide reliable and regulation-compliant autonomous business services, the research found that Zero-Touch Customer Experience systems needed a balance between governance, explainability, resilience, autonomy, and human supervision [12].

The primary motivations for interacting with retail banking AI chatbots are examined in that research. Three hundred as well as fifty retail banking clients provided information by filling out a standardized questionnaire. To discover connections and make educated guesses about the product's usage, they used machine learning and structural equation modelling (SEM). According to the results, consumers are more inclined to utilize chatbots if they see them as helpful, user-friendly, reliable, and quick to respond. The two interdependent factors of chatbot response along with adoption are linked by trust. Using Random Forest classification, the model was able to provide predictions with an accuracy of 89%. Based on these findings, financial organizations like banks may enhance and increase their usage of chatbots [13].

The study investigates the relationship between consumer happiness, trust, and loyalty on Indonesian e-commerce platforms, as well as the effects of system quality, information quality, along with service quality on satisfaction. That study uses the DeLone and McLean paradigm as its foundation. A purposive sample strategy was used to obtain creative information gathered from 453 active e-commerce customers in Indonesia. The data was acquired using Google Forms. They used PLS-SEM,

or Partial Least Squares Structural Equation Modelling, to do the study. The study's results demonstrate that high-quality systems, information, and services considerably boost consumer satisfaction. Customer loyalty is positively impacted by trust and customer pleasure. Therefore, customer loyalty is greatly impacted by an intuitive, trustworthy, and well-designed interface. To thrive in the cutthroat Indonesian market, e-commerce platforms must prioritize user interface and user experience quality while simultaneously establishing trust with their customers. The expanding function of AI in bolstering FinTech services, which are becoming more and more incorporated into e-commerce platforms, is another key finding of that research [14].

A digital twin is an avatar that captures the essence of a person and may be managed by either AI or a human. While companies have experimented with employee digital twins (e.g., virtual physicians fielding patient enquiries), the most popular use case is "celebrity-as-a-service," in which digital twins of famous people engage with fans. These digital-twin agents engage individuals in direct conversation, as opposed to the more conventional use of celebrity endorsements in promotional films. The look of a digital twin as a certain person might impact how consumers perceive it, even if it does not improve the skills of the underlying AI or human agents. Regardless of if it was operated by artificial intelligence or humans, their two trials showed that users were more engaged with digital twins of celebrities because they seemed more competent, kind, and trustworthy. It seems that celebrity impacts may protect against algorithm aversion, as the similarity of the celebrity nevertheless decreased unfavorable emotions even when mistakes occurred. Despite having the same fundamental performance, digital celebrity twins have the same impact on user trust and service uptake, making them commercially valuable in the same way as conventional celebrity endorsements [15].

The article delves at the ways in which generative AI that can work across languages might increase customer assistance and digital trust in today's online service ecosystems. Users often encounter technological complexity that limits accessibility and puts a strain on support staff as more and more public and commercial organizations use digital platforms. Conversational agents that are dependable, multilingual, and aware of their surroundings may now provide ongoing support, thanks to recent developments in big language models. Their proposal is an AI-powered customer support system that incorporates generative design principles. Its goal is to streamline user interactions, lessen the burden on support staff, and boost trust in digital services. An organized approach from issue understanding to deployment is ensured by developing the system using the CRISP-DM methodology. That study shows that a well-designed generative AI assistant can update the user experience, facilitate safe use of digital services, and meet the

increasing need for reliable and effective online help [17].

AI-Enabled Banking, Customer Loyalty, Infrastructure, and Human-AI Interaction

To investigate that dynamic environment, that study integrates literature evaluations, theoretical underpinnings, case studies, and extensive debates. Based on ideas such as Expectancy-Disconfirmation Theory (EDT) as well as Relationship Marketing Theory, the review emphasizes the significant relationship between E-Loyalty, AI, customer happiness, and loyalty. Integrating AI is changing the banking industry and making customers' experiences better, while digital E-Loyalty initiatives put an emphasis on building connections with customers that are unique, trustworthy, and value-driven. The offered theoretical framework lays the groundwork for comprehending how E-Loyalty methods powered by AI might potentially increase customer happiness and loyalty via going above and beyond, cultivating connections, improving service quality, and cultivating trust. Applying AI-driven E-Loyalty to improve client experiences, establish trust, and guarantee satisfaction is crucial for the banking business. Offering insightful findings for practitioners with future studies, that study delves deeply into the changing dynamics of customer pleasure and loyalty in banking via AI-Driven E-Loyalty initiatives [16]. Cloud computing, distributed computing, grid computing, and parallel computing are just a few examples of how the internet has revolutionized the computer industry in recent years. One of the most significant developments in computing history may have occurred in the last few years, with the rise of cloud computing. Cloud computing has numerous advantages, but many financial institutions are still unwilling to use it. The financial sector will be most affected by emerging technology, particularly cloud computing and artificial intelligence. Cloud computing allows financial institutions to react quickly to shifting market circumstances, improve client experience, and increase operational efficiency via the use of data and analytics. Therefore, cloud computing steps in to address these issues, transforming banking into a trustworthy and dependable service. That study delves into the strategy and influence of cloud computing on banking and financial institutions, as well as the substantial dependence on that technology [21].

The focus of that research is on voice assistants powered by artificial intelligence that interact with humans via natural language processing and their dataset is based on 200 recorded conversations between humans and AI-powered voice assistants from a top Korean telecom company, and they look into the complex conversational patterns that come up in these encounters. Through the use of conversation analysis techniques, with a particular emphasis on adjacency pairs, first pair-part, and second pair-part, their investigation sheds light on the ways in which AI agents as well as human users negotiate meaning or interactional roles. From the users'

SPP, they extract four different kinds of responses, which show different patterns of engagement. According to the results, users often employ keywords in response to AI-initiated prompts, which shows that they are trying to obtain information effectively. They often point out when there was no voice answer. Further evidence that cultural and language norms impact the dynamics of human-AI interaction is the usage of honorifics by Korean AI voice assistants. That highlights the importance of AI systems being able to successfully traverse social hierarchies. In the ever-changing realm of artificial intelligence-based communication, their work highlights the significance of improving human-AI discourse and offers useful implications for research and practice across disciplines [22].

Literature Summary: Past studies have concluded that customer service with AI has the potential to increase efficiency, responsiveness and personalization, however

these do not necessarily build customer trust. However, the research conducted by Trustworthy-AI has already set out some key principles on transparency, fairness, privacy, reliability, security, accountability and human oversight in this literature, which are largely centered around governance, system design, and responsible deployment. In contrast, research on customer-service has looked at trust, satisfaction, adoption and loyalty, but has tended to focus on technological or service-quality factors, rather than on all the characteristics of trustworthy-AI. However, there is limited empirical knowledge of the translation of the customer-experienced attributes of trustworthiness-AI to perceived trustworthiness-AI, the impact of this perception on perceived risk and customer trust, and the impact of trust on customer satisfaction, continuance intention and customer loyalty.

Proposed Work

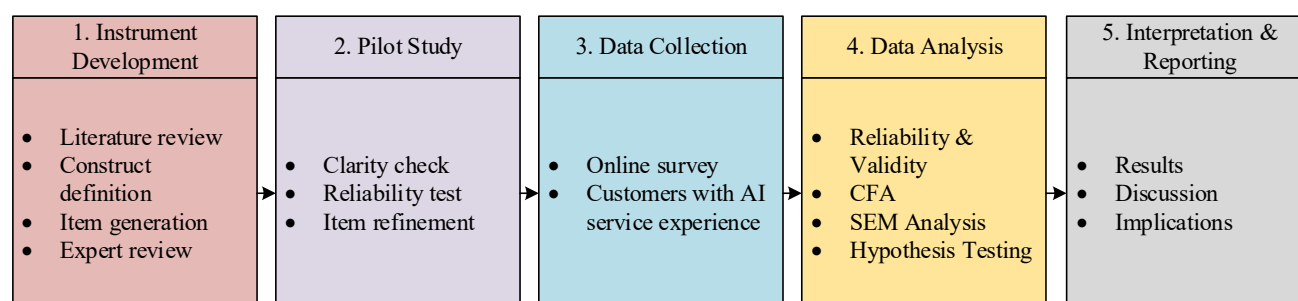


Fig. 1: Proposed Flowchart

Trust Theory

The study is based on the main theory of trust theory. Trust is an attitude of vulnerability expressed by a customer, with an expectation of a competent, honest, dependable, and good-hearted other party. Unlike human customer service, AI-powered customer service does not allow customers to see what goes on inside the AI models. They thus rely on observable cues to determine if the system is to be trusted. Transparency, reliability, privacy protection, fairness, security, human oversight and accountability serve as such signals. The study separates out AI trustworthiness perception from customer trust. Perceived AI trustworthiness is the customer's judgement on whether the AI system has trustworthy attributes, while customer trust is the customer's confidence in the AI system.

Perceived Risk Theory

Perceived risk is the reason customers might not be willing to use AI even if they know how beneficial it is. The interactions of AI can lead to a sense of privacy risk, security risk, performance risk, financial risk, and social/psychological risk. Expectations of trustworthy AI features should minimize uncertainty and vulnerabilities. Thus, perceived risk is added as a mediator between perceived AI trustworthiness and customer trust.

Technology Acceptance Perspective

It can be seen that the views on technology acceptance have provided a perspective that usefulness, ease of use, attitudes, and behavioral intentions determine technology adoption. But there is the added uncertainty of AI because customers may not know how an AI system produces outputs or handles the information. In this regard, the current study builds on the technology acceptance theories by focusing on trustworthiness and risk as key factors for the acceptance of AI in customer services.

- **Perceived AI Trustworthiness vs. Customer Trust:** It is considered as independent variable/predictor, Customer Trust as dependent variable/outcome. Customer Trust and AI Trustworthiness are not synonymous and are distinct constructs.
- **Human Oversight vs Customer Control:** It's also conceptually different. Customer Control is how much the customer feels they can influence, modify, refuse, or override AI-driven decisions, and Human Oversight is the degree of oversight over, review of, or intervention in AI decisions by humans.

Research Design

The proposed framework of trustworthy-AI is tested empirically using a quantitative and cross-

sectional research design. The quantitative approach is suitable because it is a study that investigates the relationships between a number of latent constructs and attempts to determine the statistical evaluation of the hypotheses to be tested. The research is deductive in which constructs are generated based on theories previously established and trustworthy-AI principles and then tested on primary survey data. The unit of analysis is the individual customer with experience using an AI-powered customer-service system. The aim of the Empirical Model is to explore direct and indirect relationships between dimensions of trustworthy-AI, perceived AI trustworthiness, perceived risk, customer trust, customer satisfaction, continuance intention, and customer loyalty.

Target Population

The target population is the set of customers that have experience with customer-service systems that use AI. Examples of eligible participants include individuals who have interacted with AI chatbots or virtual assistants, automated customer support applications, AI recommendation systems, AI banking assistants, AI telecom assistants, e-commerce service

assistants, insurance service assistants, travel service assistants or generative-AI customer support systems. The respondents is experienced in customer service where AI is being implemented to assess the constructs under consideration. Thus, a screening question is required at the first of the questionnaire to confirm the previous interaction with the AI customer service.

Sampling Technique and Sample Size

Because the respondents need to have first-hand experience of AI-powered customer service, a purposive sampling method can be used. The suggested sample size is 400-500 people. The sample size is suitable for estimating a structural equation model with multiple latent constructs, which is confirmed by an a priori statistical power analysis. Data is gathered from various age groups, occupations, educational attainment, income levels, and from various types of service providing industries in order to achieve better representativeness. Respondents should also vary in their level of frequency of use of AI services. The actual number of questionnaires distributed, returned, valid questionnaires, the response rate and the reasons for exclusion is reported in the final study.

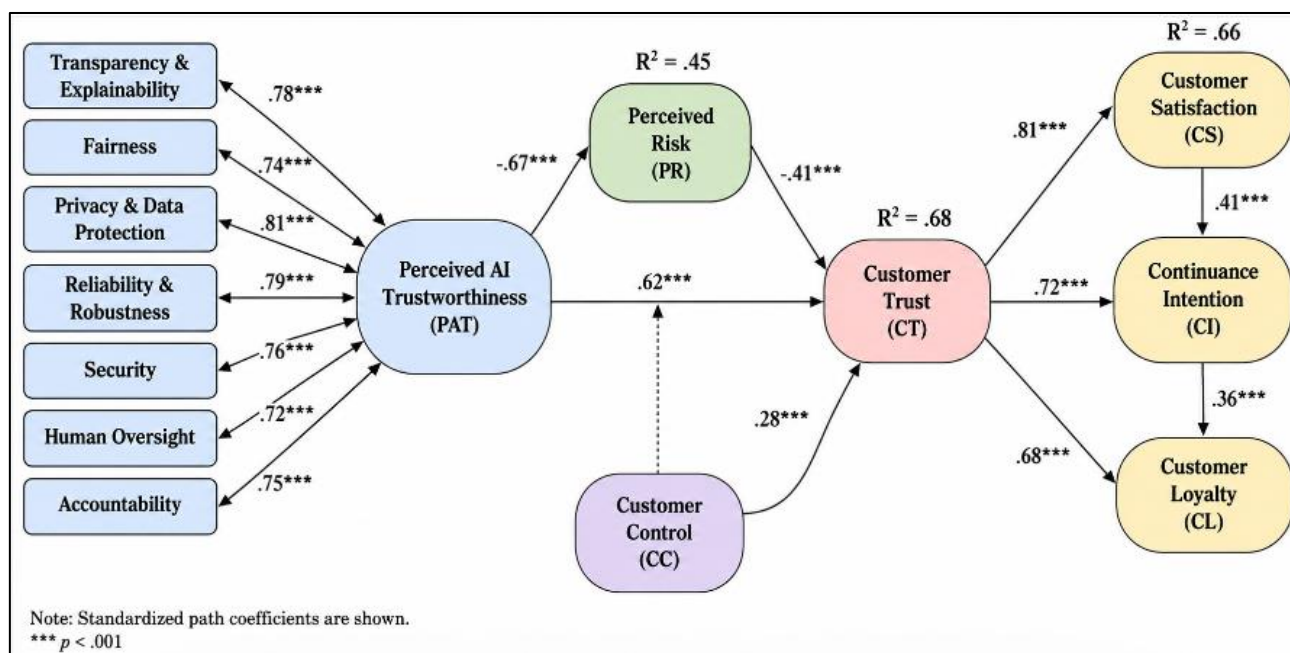


Fig. 2: SEM Analysis

Perceived Risk is defined as a mediating variable between perceived AI trustworthiness and customer trust. Customer Control is envisioned as a direct driver of customer trust, indicating the degree of control customers believe they have over or can influence or change AI decisions. Human Oversight is represented as an independent antecedent that indicates the extent to which humans oversee and intervene in the decision-making process involving AI. Therefore, without an additional moderation hypothesis and interaction term in the model, Perceived Risk is included in the model only as a mediator, and Customer Control

is included only as a direct predictor, as shown in Figure 2.

Data Collection Procedure

The primary data is collected by a structured questionnaire on self-administration. The questionnaire can be sent by online survey tools as well as by direct customer recruitment, if applicable. A pilot study is carried out prior to the main survey, with a sample of 30-50 people who have the characteristics of the desired population. The pilot study should assess the clarity of the items, length of the questionnaire, and wording of

items, construct reliability, and possible ambiguity of items. After refinement, the final questionnaire is sent to the target population. It is imperative that everyone does not feel compelled to participate. It is important that the respondents be provided with information about the purpose of the study, confidentiality of responses, their voluntary participation, and their right to withdraw from the study. All information is gathered anonymously, unless there is a necessity for and ethical approval to collect the information.

DISCUSSION

The suggested framework illustrates that customer trust in AI-powered customer service is a multi-faceted concept. While AI accuracy is crucial, it's only a component of the customer's evaluation of trust. Customers also assess the transparency, fairness, privacy protection, security, accountability and human oversight of AI. Including perceived risk is an important explanation for the growth of customer trust. The trustworthy and stabilizing nature of AI features should lead to reduced uncertainty and vulnerability, ultimately boosting confidence in using AI. It is said that the relationship implies that organizations should avoid trying to build customer trust in them through communication campaigns. There is tangible technical and organizational processes that help build trust. Customer control is also a main priority in the framework. When customers feel, they are forced to interact with AI systems, it can lead to negative experiences. Incorporating provisions for human intervention, error-correcting features, data management and the ability to question the AI's work can enhance a sense of autonomy. The proposed model then moves the focus from the use of AI to its reliance. The proposed model thus brings the focus from using AI to relying on AI. It shouldn't be about convincing customers to accept AI, despite its constraints. Instead, customers is provided enough information and protections by trustworthy AI to build the proper level of trust. In the case of generative AI, this is critical to understand. Customers may be misled into thinking that an AI system is more capable of

handling natural-language fluency than is actually the case. Thus, AI is honest about its uncertainties and limitations when necessary. When AI is unable to consistently solve a customer issue, human oversight and accountability is in place to ensure these are addressed.

Theoretical Contributions

The study's theoretical contribution is the synthesis of the concepts of trustworthy-AI and customer trust theory and perceived risk theory. The framework is not about the technical aspects of trustworthy artificial intelligence, but rather an experience-based construct for the customer. The study also separates various perceptions of AI trustworthiness from customer trust. This is significant because a system can have trustworthy features, but not be trusted by all customers. These characteristics and past interactions lead to the customers' perception of trust. Adding perceived risk as an extra mechanism of explanation. Reducing uncertainty and vulnerability is suggested to be one of the ways in which trustworthy AI can affect customer trust. The framework also brings customer-service research one step further by linking trustworthy AI with subsequent customer behavior, including satisfaction, continuance intention, and loyalty.

Questionnaire Development

There are three main parts of the questionnaire. The first section should gather information about the demographic and AI-use, such as age group, gender, education, occupation, frequency of usage of AI services, type of AI customer support platform used, the duration of using AI services, and the industry the services are used for. The second section should include the characteristics of trustworthy-AI. Transparency and explainability, fairness, privacy and data protection, reliability and robustness, security, human oversight and accountability is considered. The third section should include items that assess perceived AI trustworthiness, perceived risk, perceived customer control, customer trust, customer satisfaction, continuance intention, and customer loyalty. A five-point Likert scale could be used (1 – strongly disagree, 5 – strongly agree).

Table 1: Measurement Constructs

Construct	Code	Measurement focus
Transparency and Explainability	TE	AI disclosure, understandable explanations, clarity of limitations
Fairness	FA	Equal treatment, non-discrimination, consistency
Privacy and Data Protection	PD	Data protection, appropriate collection, privacy control
Reliability and Robustness	RR	Accuracy, consistency, dependable responses
Security	SE	Protection from unauthorized access and misuse
Human Oversight	HO	Human escalation, intervention, correction
Accountability	AC	Responsibility, complaint mechanisms, correction
Perceived AI Trustworthiness	PAT	Overall perception that AI is worthy of reliance
Perceived Risk	PR	Privacy, security, performance and uncertainty risks
Customer Control	CC	Ability to control, challenge, or terminate AI interaction
Customer Trust	CT	Confidence, willingness to rely, dependable expectations
Customer Satisfaction	CS	Overall satisfaction with AI service
Continuance Intention	CI	Intention to continue using AI service
Customer Loyalty	CL	Recommendation, preference, repeat-use intention

Table 2: Section A: Respondent Profile

S. No.	Variable	Response Options
1	Age	Below 25 / 25–34 / 35–44 / 45–54 / 55 and above
2	Gender	Male / Female / Other / Prefer not to say
3	Educational Qualification	Higher Secondary / Diploma / Undergraduate / Postgraduate / Doctorate
4	Occupation	Student / Private Employee / Government Employee / Business / Professional / Other
5	Monthly Income	Below ₹25,000 / ₹25,001–₹50,000 / ₹50,001–₹75,000 / ₹75,001–₹1,00,000 / Above ₹1,00,000
6	Type of AI Customer Service Used	Chatbot / Virtual Assistant / Generative AI Assistant / Automated Voice Assistant / Recommendation System / Other
7	Industry in Which AI Service Was Used	Banking / Retail / E-commerce / Telecommunications / Healthcare / Insurance / Travel / Other
8	Frequency of AI Customer-Service Usage	Rarely / Occasionally / Monthly / Weekly / Daily
9	Duration of AI-Service Usage	Less than 6 months / 6–12 months / 1–2 years / 2–3 years / More than 3 years
10	Previous Experience with Human Customer Service	Yes / No

Section B: Trustworthy AI Characteristics

Instruction: Please indicate your level of agreement with each statement.

Scale: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree.

Table 3: Transparency and Explainability

Code	Statement	1	2	3	4	5
TE1	The AI system clearly identifies that I am interacting with artificial intelligence.	☐	☐	☐	☐	☐
TE2	The AI system clearly explains its capabilities and limitations.	☐	☐	☐	☐	☐
TE3	The AI system provides understandable explanations for important decisions or recommendations.	☐	☐	☐	☐	☐
TE4	The information provided about the AI system is sufficiently clear.	☐	☐	☐	☐	☐
TE5	I understand how the AI system uses my information during the service interaction.	☐	☐	☐	☐	☐

Table 4: Fairness

Code	Statement	1	2	3	4	5
FA1	The AI system treats customers fairly.	☐	☐	☐	☐	☐
FA2	The AI system provides consistent service to different customers.	☐	☐	☐	☐	☐
FA3	The AI system avoids unjustified discrimination.	☐	☐	☐	☐	☐
FA4	I believe that the AI system makes service decisions impartially.	☐	☐	☐	☐	☐
FA5	The AI system provides equal opportunities for customers to receive appropriate service.	☐	☐	☐	☐	☐

Table 5: Privacy and Data Protection

Code	Statement	1	2	3	4	5
PD1	I believe that my personal information is adequately protected by the AI service.	☐	☐	☐	☐	☐
PD2	The organization clearly explains how my information is used.	☐	☐	☐	☐	☐
PD3	The AI service collects only information necessary for providing the service.	☐	☐	☐	☐	☐
PD4	I feel confident that my personal information will not be misused.	☐	☐	☐	☐	☐
PD5	I have sufficient information about the organization's data-protection practices.	☐	☐	☐	☐	☐

Table 6: B4. Reliability and Robustness

Code	Statement	1	2	3	4	5
RR1	The AI system provides accurate responses.	☐	☐	☐	☐	☐
RR2	The AI system provides consistent responses to similar questions.	☐	☐	☐	☐	☐
RR3	The AI system performs reliably during customer-service interactions.	☐	☐	☐	☐	☐
RR4	The AI system appropriately handles unexpected or complex questions.	☐	☐	☐	☐	☐
RR5	The AI system provides dependable service without frequent technical failures.	☐	☐	☐	☐	☐

Table 7: B5. Security

Code	Statement	1	2	3	4	5
SE1	I believe that the AI service protects my information from unauthorized access.	☐	☐	☐	☐	☐
SE2	I feel secure when communicating with the AI system.	☐	☐	☐	☐	☐
SE3	The AI system appears resistant to unauthorized manipulation.	☐	☐	☐	☐	☐
SE4	I trust the organization to maintain the security of the AI service.	☐	☐	☐	☐	☐
SE5	I believe that the AI service has adequate safeguards against security threats.	☐	☐	☐	☐	☐

Table 8: B6. Human Oversight

Code	Statement	1	2	3	4	5
HO1	I can request human assistance when the AI cannot resolve my problem.	☐	☐	☐	☐	☐
HO2	Human employees can intervene when necessary.	☐	☐	☐	☐	☐
HO3	I can obtain human review of important AI-related decisions.	☐	☐	☐	☐	☐
HO4	The availability of human support makes me more comfortable using AI.	☐	☐	☐	☐	☐
HO5	I can easily transfer an AI interaction to a human employee when required.	☐	☐	☐	☐	☐

Table 9: B7. Accountability

Code	Statement	1	2	3	4	5
AC1	The organization accepts responsibility for problems caused by its AI service.	☐	☐	☐	☐	☐
AC2	I can report incorrect AI responses or decisions.	☐	☐	☐	☐	☐
AC3	There is a clear mechanism for correcting AI-related errors.	☐	☐	☐	☐	☐
AC4	The organization provides appropriate support when AI produces an incorrect outcome.	☐	☐	☐	☐	☐
AC5	The organization is accountable for the consequences of its AI system.	☐	☐	☐	☐	☐

Table 10: Section C: Perceived AI Trustworthiness

Code	Statement	1	2	3	4	5
PAT1	Overall, I consider this AI service trustworthy.	☐	☐	☐	☐	☐
PAT2	I believe that this AI system operates in a dependable manner.	☐	☐	☐	☐	☐
PAT3	I consider the AI service worthy of my confidence.	☐	☐	☐	☐	☐
PAT4	I believe that this AI system can be relied upon when providing customer service.	☐	☐	☐	☐	☐
PAT5	I believe that this AI system operates responsibly.	☐	☐	☐	☐	☐

Table 11: Section D: Perceived Risk

Code	Statement	1	2	3	4	5
PR1	I am concerned that the AI system may misuse my personal information.	☐	☐	☐	☐	☐
PR2	I am concerned that the AI system may provide incorrect information.	☐	☐	☐	☐	☐
PR3	I feel uncertain about relying on AI for important customer-service issues.	☐	☐	☐	☐	☐
PR4	Using AI customer service involves potential risks for me.	☐	☐	☐	☐	☐
PR5	I worry that an AI error could negatively affect my service experience.	☐	☐	☐	☐	☐

Table 12: Section E: Perceived Customer Control

Code	Statement	1	2	3	4	5
CC1	I can control how I interact with the AI system.	☐	☐	☐	☐	☐
CC2	I can request human assistance whenever necessary.	☐	☐	☐	☐	☐
CC3	I can correct or challenge information provided by the AI.	☐	☐	☐	☐	☐
CC4	I have sufficient control over my personal information when using AI service.	☐	☐	☐	☐	☐
CC5	I can stop using the AI service when I am not satisfied with the interaction.	☐	☐	☐	☐	☐

Table 13: Section F: Customer Trust

Code	Statement	1	2	3	4	5
CT1	I trust the AI system to provide appropriate customer service.	☐	☐	☐	☐	☐
CT2	I feel confident relying on this AI service.	☐	☐	☐	☐	☐
CT3	I believe that the AI system will act consistently with my service expectations.	☐	☐	☐	☐	☐

Code	Statement	1	2	3	4	5
CT4	I am willing to depend on this AI system for customer-service assistance.	☐	☐	☐	☐	☐
CT5	I believe that the AI system has my best interests in mind when providing service.	☐	☐	☐	☐	☐

Table 14: Section G: Customer Satisfaction

Code	Statement	1	2	3	4	5
CS1	I am satisfied with my experience using AI customer service.	☐	☐	☐	☐	☐
CS2	The AI service meets my expectations.	☐	☐	☐	☐	☐
CS3	The AI service provides a satisfactory customer experience.	☐	☐	☐	☐	☐
CS4	Overall, I am pleased with the AI-enabled customer service.	☐	☐	☐	☐	☐
CS5	The AI service provides good value for my time and effort.	☐	☐	☐	☐	☐

Table 15: Section H: Continuance Intention

Code	Statement	1	2	3	4	5
CI1	I intend to continue using AI-enabled customer service.	☐	☐	☐	☐	☐
CI2	I will use AI customer service again when necessary.	☐	☐	☐	☐	☐
CI3	I prefer to continue using AI service for suitable customer-service issues.	☐	☐	☐	☐	☐
CI4	I expect my use of AI customer service to increase in the future.	☐	☐	☐	☐	☐
CI5	I would choose AI customer service again when it is available.	☐	☐	☐	☐	☐

Table 16: Section I: Customer Loyalty

Code	Statement	1	2	3	4	5
CL1	I would continue using services from an organization that provides trustworthy AI customer service.	☐	☐	☐	☐	☐
CL2	I would recommend the organization's AI-enabled service to others.	☐	☐	☐	☐	☐
CL3	Trustworthy AI customer service would strengthen my preference for the organization.	☐	☐	☐	☐	☐
CL4	I would be less likely to switch to another organization because of a trustworthy AI service experience.	☐	☐	☐	☐	☐
CL5	I would choose this organization again because of its trustworthy AI customer service.	☐	☐	☐	☐	☐

Before the main study, a pilot study is planned to help determine if the items on the questionnaire are understood by respondents. Approximately 30–50 respondents can participate in the pilot study. Cronbach's alpha is used when analyzing the pilot data. The items that show ambiguity and low factor loading or items that are redundant is checked. Next, the wording of the questionnaire is adjusted, if needed. If the research design and ethical approval of the study actually allow it, the pilot study should not be included in the main data set.

CONCLUSION

This study aims to create an empirical customer-level model that explains how trustworthy AI affects customer responses in AI-powered customer service. The updated framework defines Trustworthy AI as a multi-dimensional phenomenon, different from Perceived AI Trustworthiness which is a customer's overall evaluative judgment and Customer Trust, which refers to a psychological state that involves relying. Perceived Risk acts as a mediator that links AI trustworthiness perceptions to customer trust, and Perceived Customer Control is proposed as a boundary condition that potentially enhances the relationship between AI trustworthiness perceptions and willingness

to trust AI. Customer Trust, in turn, is then correlated with Customer Satisfaction, Continuance Intention and Customer Loyalty. A key contribution of the study is not the use of entirely new constructs, however, but the application of existing concepts in the context of AI in customer service. To address the gap between system-level research on trustworthy-AI governance and customer-level research on behavior, the framework highlights how perceived risk and customer perceptions of the observable characteristics of an AI solution may impact trust in the AI solution. The difference between Human Oversight and Perceived Customer Control also acknowledges that an organization's ability to intervene, or human oversight, differs from the customer's perception of personal control. The proposed model is empirical and needs to be validated based on survey data and using structural equation modeling. It will depend on its eventual success in developing satisfactory reliability, convergent validity, discriminant validity and structural relationships. If constructs cannot be differentiated empirically, then the model is better simplified than to keep two variables that are overlapping. It is therefore recommended that future empirical testing be conducted on the differences between Perceived AI Trustworthiness and Customer Trust, and between Human Oversight and Perceived Customer Control.

REFERENCES

- Feng, Y. (2025). Integrating AI-Based Chatbots for Automated Customer Support in E-Commerce: Using NLP Models like GPT for Real-Time Customer Service Automation. *Science and Technology of Engineering, Chemistry and Environmental Protection*.
- Shih, Y., & Kang, Y. (2025). Design and Implementation of a Secure RAG-Enhanced AI Chatbot for Smart Tourism Customer Service: Defending Against Prompt Injection Attacks - A Case Study of Hsinchu, Taiwan. *ArXiv*, abs/2509.21367.
- Reddy, D.P. (2024). How do customers react to the integration of AI in customer experience. *International Journal For Multidisciplinary Research*.
- Birru, N.K. (2025). Secure AI Infrastructure: Building Trustworthy AI Systems in Distributed Environments. *World Journal of Advanced Engineering Technology and Sciences*.
- Chawla, M., Mukherjee, A., Prasad, L., Shetty, N., & Sharma, R. (2024). Toward Trustworthy AI Systems: A Converged Architecture for Governance, Reliability, and Automated Testing in Enterprise Platforms. *International Journal of Emerging Trends in Computer Science and Information Technology*.
- Ed-Daakouri, I., Alla, L., Nfaoui, E.H., & Elhissoufi, M. (2025). Generative AI in Smart Customer Experience Management: A Bibliometric Insight. *2025 IEEE International Conference on Advanced Technologies in Supply Chain Management (ATSCM)*, 1-6.
- Yuvaraj, N., & Kumar, M.S. (2023). Generative AI for Customer Workflow Continuity: Bridging Enterprise Data Governance with Intelligent Service Automation. *American International Journal of Computer Science and Technology*.
- Fürst, A., Trißler, L., Friedrich, R., & Wirtz, J. (2025). Service recovery by AI or human agents: Do failure and strategy context matter? *Journal of Service Management*.
- Bianchi, L., Romano, G., & Conti, F. (2024). Governed Cloud AI for Financial Services: MLOps, Foundation Models, and Trustworthy AI Analytics across AWS, Azure, and Google Cloud. *Journal of AI Analytics and Applications*.
- Kshetri, N., & Kshetri, N. (2025). The Rise of Agentic AI in Finance: Opportunities, Risks, and Human-Centric Integration. *IT Professional*, 27, 19-24.
- Chahar, S., Kaswan, K.S., Sharma, M., & Dhatwal, J.S. (2025). Trustworthy AI for Smart Community at Tourist Destinations. *2025 International Conference on Sustainability, Innovation & Technology (ICSIT)*, 1-6.
- Yuvaraj, N., & Kumar, M.S. (2025). Agentic AI for Zero-Touch Customer Experience: A Governance-Constrained Framework for Autonomous Service Systems. *American International Journal of Computer Science and Technology*.
- A, V., & G, T. (2025). Analyzing Customer Adoption of AI-Based Chatbots in Retail Banking Using a Hybrid Structural Equation Modeling Approach. *2025 3rd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, 1, 627-630.
- Abigael, V., Tafida, D., Yumna, Kania, M., & Nafilah, Z.F. (2025). Building Customer Loyalty in E-Commerce with User Experience and Interface Design While Adapting to AI Developments in FinTech. *2025 International Conference on ICT for Smart Society (ICISS)*, 1-6.
- Yuan, L.(., Kim, A., Seymour, M., & Dennis, A.R. (2026). Celebrity-as-a-Service: Trust and Willingness to Use Digital Twins for Customer Service. *Journal of Management Information Systems*, 43, 335 - 367.
- James, B.V., Joseph, D., & Sharma, T. (2024). Transforming Banking Services: AI-Driven E-Loyalty Strategies and Case Study Insights on Customer Satisfaction and Loyalty Enhancement. *Proceedings of the 1st International Conference on Artificial Intelligence, Communication, IoT, Data Engineering and Security, IACIDS 2023, 23-25 November 2023, Lavasa, Pune, India*.
- Slim, M.A., Chaibi, M.T., & Abdelkrim, A. (2026). Deploying a Multilingual Generative AI System for Customer Support: A CRISP-DM Approach. *2026 IEEE International conference on Advanced Systems and Emergent Technologies (IC_ASET)*, 1-6.
- Vatankhah, S., Bamshad, V., Arıcı, H.E., & Duan, Y. (2024). Ethical implementation of artificial intelligence in the service industries. *The Service Industries Journal*, 44, 661 - 685.
- Li, Z. (2024). AI Ethics and Transparency in Operations Management: How Governance Mechanisms Can Reduce Data Bias and Privacy Risks. *Journal of Applied Economics and Policy Studies*.
- Gantumur, Khuslen. (2025). The Impact of Artificial Intelligence on Online Shopping. *European Scientific Journal*, ESJ. 21. 1. 10.19044/esj.2025.v21n4p1.
- Parne, P. (2021). Cloud Computing Strategy and Impact in Banking/Financial Services. *Computer Science and Information Technology Trends*.
- Ahn, S., Kim, S., Lee, E., Lee, H., Song, E., Song, J., & Won, J. (2024). How do AI and human users interact? Positioning of AI and human users in customer service. *Text & Talk*, 45, 301 - 318.

Cite This Article: Patrick Monis & Peeyush Kumar Pandey (2026). Trustworthy AI in Customer Service: A Conceptual Framework for Building Customer Trust. *East African Scholars J Eng Comput Sci*, 9(2), 48-58.
