Is Auditive Communication with ChatGPT an Effective Means of Building Trust Between People and Machines: A Quantitative Study

Marvin Tessitore
Technical University of Applied Sciences Wuerzburg-Schweinfurt
Faculty of Computer Science and Business Informatics
Wuerzburg, Germany
email: Marvin.tessitore.doktoranden@thws.de

Hakan Arda
Technical University of Applied Sciences Wuerzburg-Schweinfurt
Faculty of Computer Science and Business Informatics
Wuerzburg, Germany
email: Hakan.Arda.doktoranden@thws.de

Nicholas Müller
Technical University of Applied Sciences Wuerzburg-Schweinfurt
Faculty of Computer Science and Business Informatics
Wuerzburg, Germany
email: Nicholas.mueller@thws.de

Karsten Huffstadt
Technical University of Applied Sciences Wuerzburg-Schweinfurt
Faculty of Computer Science and Business Informatics
Wuerzburg, Germany
email: Karsten.huffstadt@thws.de

***Abstract*—Rapid advances in Artificial Intelligence (AI) have accelerated the integration of conversational agents into everyday tasks. While voice-based interaction is becoming increasingly prevalent, its influence on user trust in AI systems remains insufficiently understood. Existing research has largely focused on text-based interfaces, leaving open whether auditory interaction can enhance or even diminish perceived trustworthiness. This study empirically examines whether the communication modality of ChatGPT (text vs. auditory) affects users' trust in the system. In a controlled experiment, participants with diverse backgrounds interacted with ChatGPT to complete story-based tasks requiring nuanced reasoning. Trust was measured through a nine-item quantitative questionnaire grounded in the Technology Acceptance Model (TAM). The results show that speech-based interaction was associated with significantly higher general trust in technological systems (Q1: $p = 0.019$, $d = 0.59$). No significant differences were found for perceived truthfulness, doubts about system accuracy, usefulness, or ease of use. These findings suggest that trust formation depends less on the interaction channel and more on underlying system qualities, such as accuracy, coherence, and conversational competence. The study provides new insights for designers of AI-driven voice systems: resources should be prioritised toward improving response quality and transparent system behaviour rather than assuming inherent trust benefits from auditory communication.**

Keywords - Human–AI Interaction; User Trust; Voice Interface; Technology Acceptance Model; Conversational AI.

I. INTRODUCTION

Over the past decade, AI has evolved into a central component of everyday digital interaction [1]. Large Language Models (LLMs), such as ChatGPT, in particular, have transformed expectations of conversational systems through their ability to generate coherent, context-sensitive, and human-like responses [2]. The rapid adoption of ChatGPT is illustrated in Figure 1, which shows its weekly active user base growing to over 100 million users within months of its release, underscoring the societal relevance of research into human–AI interaction. As these technologies continue to advance, understanding the human factors that influence their acceptance has become increasingly important, especially with regard to trust [3][4].

Trust is widely recognized as a key determinant of successful human–AI interaction [3][4]. Previous research indicates that trust formation is shaped by factors, such as perceived competence, transparency, contextual relevance, and users' prior attitudes toward AI systems [3][5]. Studies in domains, such as healthcare and education further suggest that interaction modality (text, speech, or multimodal) can influence perceptions of credibility and reliability [6][7][8].

Although voice assistants, such as Amazon Alexa, Google Home, and Apple Siri are widely adopted [9], the impact of auditory interaction on trust in advanced language models remains insufficiently explored [10][11]. Voice-based interfaces offer a more natural mode of communication [6][8], yet it is unclear whether they enhance, reduce, or merely replicate trust dynamics observed in text-based interaction [12].

This study addresses this gap by examining whether communication modality (text versus speech) affects users' trust in ChatGPT. In a controlled experiment, participants completed story-based tasks and evaluated the system using a structured quantitative questionnaire measuring perceived reliability and competence [13].

The aims of this research are twofold: first, to determine whether voice-based interaction influences trust differently than text-based interaction; and second, to provide empirical evidence to support the design of transparent, reliable, and user-centred conversational AI systems.

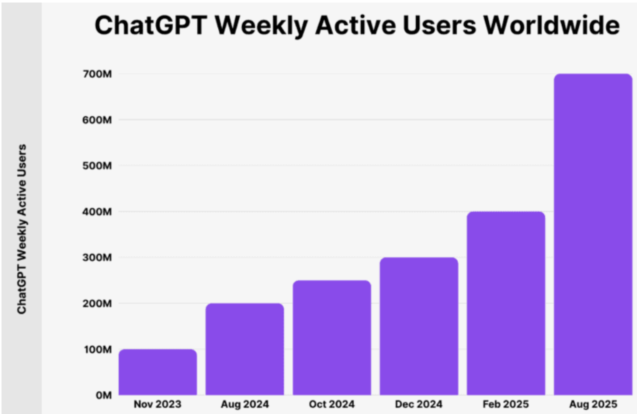

Figure 1. ChatGPT: Weekly Active Users.

The remainder of this paper is organized as follows: Section II reviews related work on trust and human–AI interaction. Section III describes the methodology and theoretical background. Section IV presents the questionnaire design. Section V outlines the experimental setup and participants. Section VI reports the results. Section VII discusses the findings in relation to the research hypotheses. Section VIII presents conclusions and directions for future work.

II. Related Works

A. *Trust and Risk in a Digital World*

The report "Trust and Risk in a Digital World" [14] examines key mechanisms of trust formation in virtual teams and online environments based on extensive empirical evidence. It identifies several determinants that influence how individuals establish and maintain trust in digitally mediated contexts.

A central finding highlights the restorative role of face-to-face interaction when trust is compromised in electronic communication. This emphasizes that, despite increasing digitalization, direct interpersonal contact remains an important mechanism for rebuilding trust, posing challenges for systems operating exclusively in virtual environments.

Furthermore, the report underscores the influence of interface design on initial trust perceptions. Factors, such as visual aesthetics, usability, and information clarity significantly affect users' evaluations of credibility and reliability, suggesting that trust is shaped not only by social factors but also by design characteristics.

The authors also emphasize the growing importance of the internet in everyday life, arguing that trust considerations must be systematically integrated into digital services that function as primary sources of information and communication.

Although the report does not explicitly address artificial intelligence, its findings are transferable to AI-driven systems. Since conversational agents operate within digital interfaces, principles, such as transparency, clarity, and perceived reliability are likely to influence trust in human–AI interaction.

Overall, this work provides a theoretical foundation for understanding trust in virtual environments and offers relevant implications for the design of trustworthy AI interfaces.

B. *Improved Trust in Human–Robot Collaboration with ChatGPT*

The study "Improved Trust in Human–Robot Collaboration with ChatGPT" [15] investigates the use of ChatGPT as a conversational control interface in a Human–Robot Collaboration (HRC) setting. The authors analyze how natural language interaction influences operator performance and trust in robotic systems.

In a controlled experiment with 15 participants, users assembled a workpiece using a ChatGPT-based assistant that issued commands to a robotic arm. The study was conducted in a virtual reality environment to enable detailed behavioral observation. Performance was evaluated based on task completion time and self-reported ratings.

Results showed that the ChatGPT-enabled interface significantly improved task efficiency compared to conventional fixed-command systems. Participants reported that the assistant's ability to retain contextual information and adapt to prior interactions contributed to smoother and more intuitive task execution.

Trust and cognitive load were measured using standardized questionnaires. The findings indicate reduced mental effort and increased trust when participants interacted with the conversational interface. These effects were attributed to the system's natural language capabilities and its capacity for contextual continuity.

Overall, the study demonstrates that naturalistic communication with ChatGPT can enhance both performance and trust in collaborative robotic systems. While the focus lies on physical human–robot interaction, the results suggest that conversational AI can positively influence trust formation in technologically mediated environments, which is relevant to the present study.

Prior research has established that conversational interfaces can enhance user trust compared to traditional text-based interaction. Gupta et al. [22] demonstrated that dialogue-based AI agents increase perceived trust in information systems, while Rheu et al. [23] showed that conversational framing positively influences user attitudes toward automated systems. The present study extends this line of research in two important ways: first, by focusing specifically on LLM-based conversational AI (i.e., ChatGPT) rather than general-purpose chatbots or rule-based dialogue systems; and second, by comparing voice-based and text-based modalities within the same AI system using the TAM as a validated measurement framework. This allows for a direct, controlled assessment of modality effects in the context of current-generation generative AI systems, which remains an underexplored area in the existing literature.

III. Methodology and Implementation

A. Research Concept

The present study employs a quantitative research design to examine how trust is formed in interactions with ChatGPT. This methodological choice is grounded in the objective of generating an initial empirical understanding of trust-related factors in the early adoption phase of large language models [16]. Quantitative methods are particularly suited for this purpose, as they enable systematic data collection, statistical comparison between groups, and the extraction of generalizable patterns, which are important requirements for exploratory work on emerging technologies [17].

Given that ChatGPT represents a relatively new class of AI systems, a descriptive, data-driven approach is essential to map the trust landscape before more complex theoretical models or causal mechanisms can be meaningfully investigated. Early-stage research benefits from quantification, as it allows researchers to identify trends, user perceptions, and recurring evaluation patterns without relying on preconceived assumptions [18][19].

To address these aims, the study was structured around two experimental conditions representing different communication modalities. Participants were assigned to one of these groups and completed standardized tasks designed to elicit interaction with ChatGPT. Following the experiment, participants' attitudes, perceptions, and trust assessments were captured through a structured questionnaire.

The conceptual basis of this questionnaire draws on the TAM [20], a well-established framework for examining user acceptance of novel technologies. TAM focuses on perceived usefulness and perceived ease of use as central determinants of user attitudes and behavioral intentions. These constructs provide a theoretically grounded lens for assessing trust in AI systems, as trust is closely tied to perceptions of system competence, reliability, and effortlessness of interaction. By integrating TAM into the measurement approach, the study ensures that user trust is evaluated systematically and in alignment with established technology acceptance theory [20].

B. Theoretical Background

The TAM, originally introduced by Davis [20], is one of the most widely applied frameworks for understanding how individuals evaluate, accept, and use technological systems. Over several decades, TAM has demonstrated strong predictive validity across diverse technological contexts and remains a foundational model in information systems research.

At the center of TAM are two core constructs: Perceived Usefulness (PU) and Perceived Ease of Use (PEU) [20].

- Perceived Usefulness refers to the degree to which an individual believes that using a particular system enhances their performance or supports task accomplishment. In the context of ChatGPT, this includes how effectively users feel the system helps them generate information, solve tasks, or achieve specific goals.

- Perceived Ease of Use describes the extent to which an individual believes that interacting with a system requires minimal effort. Applied to ChatGPT, this relates to how intuitively users can formulate prompts, understand responses, and operate the system without cognitive strain.

TAM proposes that both constructs shape users' attitudes toward a technology, which subsequently influence their intention to use and, ultimately, their actual usage behavior [20]. As trust is closely tied to perceptions of system capability, reliability, and predictability, TAM provides a meaningful theoretical basis for examining how perceived usefulness and ease of use contribute to trust in AI-driven conversational systems, such as ChatGPT.

For the purpose of the present study, the original version of TAM was deliberately selected as the theoretical foundation. More recent extensions, such as TAM2, TAM3, and the Unified Theory of Acceptance and Use of Technology (UTAUT), introduce additional constructs including social influence, facilitating conditions, or perceived enjoyment. While these models offer broader explanatory power, their complexity may obscure the specific focus of this study: understanding how basic perceptions of usefulness and ease of use relate to user trust. Future research could build upon the current work by integrating extended TAM versions to capture more nuanced determinants of trust in AI technologies.

As shown in Figure 2, the original TAM structure proposed by Davis [20] is illustrated.

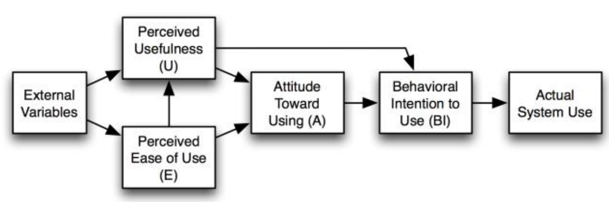

Figure 2. Technology Acceptance Model.

IV. Questionnaire

Based on the assumption that trust in technological systems is shaped by perceived usefulness and perceived ease of use [20], the questionnaire was designed to reflect the core constructs of the TAM. The items were formulated to capture participants' evaluations of ChatGPT along these dimensions while including additional questions explicitly targeting trust as an independent psychological factor.

All items were presented as questions and answered on a five-point Likert scale ranging from 1 (strongly disagree) to

5 (strongly agree). This format was chosen to enable standardized quantitative comparison between participants and experimental conditions. Prior to data collection, the questionnaire was reviewed for clarity and consistency to ensure that all items were comprehensible and unambiguous.

The questionnaire consisted of the following items:

1. Do you generally trust machines or systems in your environment?
2. Do you think you have acquired new knowledge?
3. Would you pass on what you have learned to friends or acquaintances?
4. Would you use the information you received for academic work at university or at the TH?
5. Did you feel that your questions were answered competently?
6. Did you feel sufficiently supported by the system?
7. Did you doubt the information provided by the system at any point?
8. How truthful do you consider the generated text to be?
9. Would you use the system again to acquire knowledge?

A substantial proportion of the items was oriented toward perceived usefulness (e.g., Items 2, 3, 5, and 9), as this construct reflects participants' evaluation of the system's relevance and practical value. Trust was measured through targeted items addressing both general attitudes and situational evaluations (Items 1, 7, and 8), enabling respondents to express their confidence in the system's reliability and truthfulness.

Perceived ease of use was captured primarily through Item 6, with partial relevance in Item 9. Since voice-based interaction is generally expected to reduce operational barriers, this construct was intentionally represented in a concise manner. This approach ensured that the questionnaire remained focused while still covering the central TAM dimensions relevant to the present study.

The allocation of questionnaire items to the respective TAM constructs and trust category is summarized in Table I.

TABLE I. ASSIGNMENT OF QUESTIONNAIRE ITEMS TO TAM DIMENSIONS

Q-Number	*TAM*
Q1	Trust
Q2	Usefulness
Q3	Usefulness (private)
Q4	Usefulness (research)
Q5	Usefulness
Q6	Ease of use
Q7	Trust
Q8	Trust
Q9	Ease of use and Usefulness

V. EXPERIMENTAL SETUP

A. *Test A: Auditory Interaction with ChatGPT*

For the voice-based condition, a custom prototype was developed using a laptop, the ChatGPT model, Microsoft Azure speech services, and C++ integration. Spoken input was converted to text via Azure, processed by ChatGPT, and returned to participants through text-to-speech output.

The system operated in a continuous listening mode without manual activation. An initial feedback loop, caused by the system recognizing its own speech output as input, was resolved through targeted adjustments in the C++ implementation.

All sessions were conducted in a quiet office environment to minimize background noise. Participants entered the room individually and received a printed task description identical to that of the text-based condition. Tasks required them to retrieve information on historical topics using spoken interaction only, ensuring a focused voice-based experience.

Following the interaction, participants completed a structured nine-item TAM-based questionnaire assessing perceived usefulness, ease of use, and trust.

B. *Test B: Text-Based Interaction with ChatGPT*

For the text-based condition, participants interacted with ChatGPT via the standard interface using keyboard input and received written responses, enabling a continuous text-based dialogue. All sessions were conducted in a controlled, distraction-free environment to ensure consistency and reliable data collection.

Participants entered the room individually and received a printed task description identical to that of the auditory condition, ensuring comparable task requirements and cognitive demands across modalities. The standardized task sheet minimized interface-related variability and supported a consistent experimental procedure.

Following the interaction, participants completed the same nine-item TAM-based questionnaire to assess perceived usefulness, ease of use, and trust, enabling direct comparison with the auditory condition.

C. *Participants and Procedure*

The final dataset comprised N = 66 participants, evenly distributed across the two conditions (33 per group), with a mean age of 23.51 years. Most participants were students or employees of the Technical University of Applied Sciences Würzburg-Schweinfurt (THWS).

Each session lasted approximately 5–7 minutes. All participants were instructed to ask the same number of questions and were given identical opportunities to interact with the system, ensuring comparability between the speech-based and text-based conditions.

D. *Hypothesis*

The central aim of this study is to examine whether the modality of interaction with ChatGPT, comparing speech-based versus text-based interaction, affects the level of trust users place in the system. To investigate this question empirically, the following hypotheses were formulated:

- H0 (Null Hypothesis):
  Speech-based input and output have no effect on the degree of trust users place in ChatGPT.
- H1 (Alternative Hypothesis):
  Speech-based input and output have an effect on the degree of trust users place in ChatGPT.

These hypotheses provide the basis for the comparative analysis between the auditory and text-based interaction conditions and guide the statistical evaluation performed in this study.

VI. RESULTS

To examine whether interaction modality (speech-based vs. text-based) influenced user perceptions, independent two-sample t-tests were conducted for each questionnaire item. The analysis was structured according to three categories derived from the TAM: usefulness, ease of use, and trust [20].

Usefulness reflects participants' evaluations of the system's practical value for information retrieval and learning, ease of use captures perceived interaction effort, and trust represents assessments of credibility and reliability. These categories provide a structured framework for comparing the two experimental conditions.

A. Response Distribution

Figure 3 presents box-and-whisker plots comparing the response distributions for the speech-based and text-based conditions across all nine questionnaire items. The speech-based condition (SP) is shown in blue and the text-based condition (TX) in red.

Each box represents the interquartile range (IQR), encompassing the middle 50% of responses. The horizontal line inside each box indicates the median response. Whiskers extend to the most extreme values within 1.5 times the IQR, and individual data points beyond this range are plotted as outliers.

Across most items, the two conditions show comparable medians and IQRs, indicating broadly similar response patterns. A notable exception is Q1 (general trust in the system), where the speech-based condition yields a visibly higher median and a more compact distribution, consistent with the statistically significant result ($p = 0.019$). Q7 (doubt about information) shows the most divergent distributions: the text-based group is clustered at low doubt scores, while the speech-based group displays greater spread.

Overall, the boxplots confirm that response patterns were largely equivalent across modalities, with the exception of Q1. The directional trends for Q3 and Q4 (marked †) are visible as a slight upward shift in the text-based boxes, though the overlap in IQRs reflects the non-significant p-values.

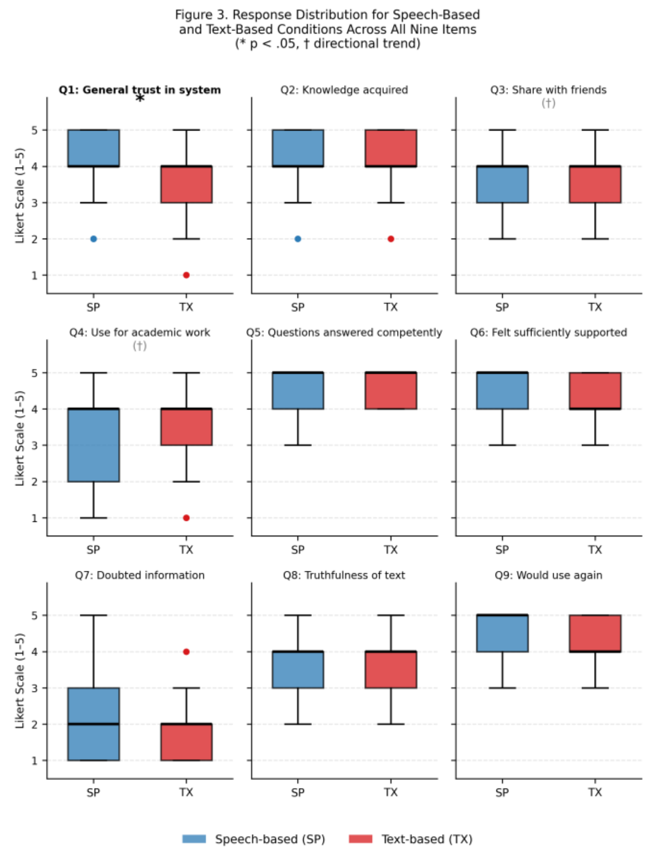

Figure 3. Response Distribution for Speech-Based and Text-Based Conditions Across All Nine Questionnaire Items.

B. Independent Two-Sample t-Test

To examine whether interaction modality (speech-based vs. text-based) influenced participant responses, independent two-sample t-tests were conducted for each questionnaire item using IBM SPSS Statistics.

Statistical analysis revealed a significant difference between the two conditions for Item Q1 only ($t(64) = 2.41$, $p = 0.019$, $d = 0.59$, 95 % CI [0.09, 0.94]), leading to the rejection of the null hypothesis for this item. Items Q3 ($p = 0.131$) and Q4 ($p = 0.066$) showed directional trends that did not reach statistical significance. All remaining items (Q2, Q5–Q9) showed no meaningful differences between conditions (all $p > 0.10$). Table II presents the group means, standard deviations, t-values, p-values, and effect sizes (Cohen's d) for all nine items.

Table II presents descriptive and inferential statistics (means, standard deviations, t-values, p-values, Cohen's d, and 95 % confidence intervals) for all nine items.

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-383-5

TABLE II. DESCRIPTIVE AND INFERENTIAL STATISTICS (N = 66, N = 33 PER CONDITION). SP = SPEECH, TX = TEXT; D = COHEN'S D; CI = 95 % CONFIDENCE INTERVAL FOR MEAN DIFFERENCE. * P < 0.05.

Item	Speech M (SD)	Text M (SD)	t(64)	p	d	95 % CI
Q1	4.00 (0.83)	3.49 (0.91)	2.41	0.019	0.59	[0.09, 0.94]
Q2	4.03 (0.95)	4.03 (0.98)	0.00	1.000	0.00	[−0.48, 0.48]
Q3	3.73 (0.84)	4.03 (0.77)	−1.53	0.131	−0.38	[−0.70, 0.09]
Q4	3.27 (1.31)	3.82 (1.04)	−1.87	0.066	−0.46	[−1.13, 0.04]
Q5	4.61 (0.50)	4.67 (0.48)	−0.51	0.615	−0.12	[−0.30, 0.18]
Q6	4.52 (0.57)	4.36 (0.60)	1.05	0.297	0.26	[−0.14, 0.44]
Q7	2.18 (0.98)	1.79 (0.82)	1.77	0.082	0.44	[−0.05, 0.84]
Q8	4.12 (0.55)	4.12 (0.65)	0.00	1.000	0.00	[−0.30, 0.30]
Q9	4.45 (0.62)	4.55 (0.51)	−0.66	0.515	−0.16	[−0.37, 0.19]

C. Reliability Analysis: Cronbach's Alpha Test

To assess the internal consistency of the questionnaire, a Cronbach's Alpha reliability analysis was conducted. The resulting coefficient of $\alpha = 0.547$ indicates a moderate level of internal consistency. Although this value falls below the commonly recommended threshold of 0.70, its interpretation must be considered in light of the conceptual structure of the instrument.

The questionnaire integrates multiple theoretically distinct constructs, including perceived usefulness, perceived ease of use, and trust. As these dimensions capture different aspects of user perception, heterogeneous response patterns are expected. Such multidimensionality typically leads to lower overall alpha values when reliability is calculated across all items, without necessarily indicating insufficient measurement quality [21].

Given the exploratory nature of the present study and its focus on capturing a broad range of user perceptions, a unified reliability analysis was considered appropriate. The moderate alpha value therefore reflects the instrument's multi-construct design rather than methodological inadequacy.

No items were removed, as each question contributes relevant information to the investigation of trust formation in human–AI interaction. Retaining all items preserves the conceptual breadth of the measurement approach and supports the study's aim of providing an initial empirical assessment of modality-dependent trust dynamics.

Reporting sub-scale reliability separately for the Trust (Q1, Q7, Q8), Usefulness (Q2, Q3, Q4, Q5, Q9), and Ease of Use (Q6, Q9) dimensions would provide a more granular assessment of measurement consistency. However, as data were collected using printed questionnaires and only aggregated frequency distributions were recorded, individual response vectors required for sub-scale Cronbach's alpha computation are not available. Future studies should ensure digital data capture to enable sub-scale reliability analysis.

VII. DISCUSSION

The present study employed an A/B testing design and an independent two-sample t-test to investigate whether the modality of interaction with ChatGPT, comparing speech-based versus text-based modalities, influences user perceptions of trust, usefulness, and ease of use. Statistical analysis revealed a significant between-group difference for one item only: Q1 (general trust in machines; $p = 0.019$). All other items did not yield statistically significant differences (all $p > 0.05$). Directional but non-significant trends were observed for Q4 ($p = 0.066$) and Q3 ($p = 0.131$), both favouring the text condition.

To interpret these findings more precisely, the questionnaire was structured into three theoretical categories derived from the TAM: Usefulness, Ease of Use, and Trust. In the following subsections, each category is discussed individually with respect to the corresponding questionnaire items. This structure allows for a more differentiated understanding of how the interaction modality affects distinct dimensions of user perception.

Finally, the discussion concludes with a synthesis across all categories. This integrative perspective highlights how usefulness, ease of use, and trust interact with one another and collectively shape user judgments in human–AI interaction. By connecting these findings, the study aims to provide a holistic interpretation of how communication modality may, or may not, influence the broader acceptance of AI systems, such as ChatGPT.

A. Category: Trust

The Trust category comprised the following questionnaire items:

- Q1: Do you generally trust machines or systems in your environment?
- Q7: Did you doubt the information provided by the system?

- Q8: How truthful do you consider the generated text to be?

Analysis using the independent two-sample t-test (see Table II) showed that only Q1 yielded a statistically significant difference between the two groups. Participants in the speech-based condition reported a higher general trust in machines and systems than those in the text-based condition. This suggests that the auditory modality may enhance users' baseline trust in technology during interaction.

In contrast, Q7 (doubt about information: $p = 0.082$, $d = 0.44$) and Q8 (perceived truthfulness: $p = 1.000$, $d = 0.00$) showed no significant differences between the groups. Notably, Q7 approached but did not cross the conventional significance threshold, suggesting a weak trend toward greater expressed doubt in the text condition ($M = 1.79$) compared to the speech condition ($M = 2.18$) that warrants investigation in larger samples.

Taken together, the findings for this category suggest that while specific trust-related evaluations (credibility or doubts) are not affected by modality, generalized trust toward the system appears to be higher when interaction occurs via speech. This indicates that the communicative channel may influence broader perceptions of trust, even if it does not alter assessments of specific system outputs.

B. *Category: Usefulness*

The Usefulness category included Items Q2, Q3, Q4, Q5, and Q9, addressing perceived knowledge acquisition, information dissemination, academic applicability, response competence, and continued system use.

Statistical analysis did not reveal significant differences between the two conditions for any usefulness item. Item Q2 (knowledge acquisition) showed no difference ($p = 1.000$, $d = 0.00$), confirming that both modalities were equally effective in supporting learning. Item Q9 (willingness to reuse the system) was also comparable across conditions ($p = 0.515$, $d = 0.16$).

Items Q3 and Q4 showed directional trends that did not reach statistical significance. Participants in the text-based condition tended to report higher intentions to share learned content with others (Q3: $M_{text} = 4.03$ vs. $M_{speech} = 3.73$; $p = 0.131$, $d = 0.38$) and to use information for academic purposes (Q4: $M_{text} = 3.82$ vs. $M_{speech} = 3.27$; $p = 0.066$, $d = 0.46$), though these differences remain inconclusive at the conventional $\alpha = 0.05$ level.

These trends are consistent with the intuitive advantage of written text for information retention and reuse: text can be scrolled back, copied, and cited, whereas spoken output is transient. Future research with larger samples should examine whether these tendencies reach statistical significance.

Item Q5 (perceived competence of responses) showed no significant difference between conditions ($p = 0.615$, $d = 0.12$), with both groups rating competence highly ($M_{speech} = 4.61$, $M_{text} = 4.67$). This suggests that the perceived quality of ChatGPT's responses was independent of the interaction modality.

Overall, the data indicate that usefulness perceptions were broadly equivalent across modalities. Neither condition produced systematically higher ratings across the usefulness dimension.

In summary, the hypothesis that modality significantly differentiates usefulness perceptions is not supported by the present data. Both interaction modes were equally effective for knowledge acquisition and willingness to reuse the system. The directional trends for Q3 and Q4 suggest a potential advantage of text for information sharing and academic use, but these require replication with larger, more diverse samples before conclusions can be drawn.

C. *Category: Ease of Use*

The Ease of Use category comprised Items Q6 and Q9, addressing perceived system support and willingness to use the system again for knowledge acquisition.

Statistical analysis showed highly similar response patterns across the speech-based and text-based conditions for both items, indicating that interaction modality had no significant effect on perceived ease of use.

This finding is noteworthy given that voice-based interaction is generally less familiar to many users than text-based input. Despite this, participants reported that the speech-based system was easy to use and did not introduce additional cognitive or technical barriers.

Both groups expressed comparable willingness to use the system again (Q9), suggesting that neither modality hindered continued engagement. While familiarity may account for positive evaluations of the text-based interface, the similarly favourable assessment of the speech-based variant indicates that participants were able to interact with the voice system without difficulty.

Overall, the results demonstrate that no meaningful differences emerged between the two conditions with respect to ease of use. Speech-based interaction was perceived as equally accessible and user-friendly as text-based communication.

D. *Integrated Summary of Findings*

The comparative analysis of the three dimensions (Trust, Usefulness, and Ease of Use) provides insights into how interaction modality shapes users' perceptions of ChatGPT.

In the Trust category, speech-based interaction was associated with significantly higher general trust in technological systems (Q1: $M_{speech} = 4.00$ vs. $M_{text} = 3.49$; $t(64) = 2.41$, $p = 0.019$, $d = 0.59$). This constitutes a medium-sized effect and represents the only statistically significant finding in the study. Perceived truthfulness (Q8) and expressed doubt (Q7) did not differ significantly, indicating that the modality effect is limited to

Copyright (c) IARIA, 2026. ISBN: 978-1-68558-383-5

a baseline or dispositional trust in the technology rather than situation-specific credibility judgements.

With respect to Usefulness, no significant differences were found across any of the five items. Both modalities equally supported knowledge acquisition (Q2) and willingness to reuse the system (Q9). Directional trends suggested that text-based output may be associated with greater intentions to share information (Q3) and apply it academically (Q4), but these differences did not reach statistical significance and should be interpreted with caution.

For Ease of Use, no meaningful differences were identified between the two conditions. Both interaction modes were perceived as equally accessible and user-friendly, and participants reported comparable levels of support and willingness to use the system again. This suggests that speech-based interaction does not introduce additional usability barriers.

Overall, the results demonstrate that interaction modality influences specific dimensions of user perception rather than overall evaluations of the system. Speech-based interaction enhances generalized trust in technological systems, whereas text-based interaction shows a non-significant directional trend toward supporting information sharing and academic applicability. Both modalities are regarded as equally easy to use.

Accordingly, the null hypothesis is rejected for Q1 only. For all other items, the data are consistent with the null hypothesis of no modality effect. Overall, the results suggest that interaction modality has a limited and specific influence on user perceptions of ChatGPT: it elevates general trust in the technology without altering assessments of content quality, usefulness, or ease of use.

VIII. CONCLUSION AND FUTURE WORK

This study provides empirical insights into interaction modality (speech versus text), which influences user perceptions of ChatGPT with respect to trust, perceived usefulness, and ease of use. Statistical analysis of a nine-item TAM-based questionnaire administered to $N = 66$ participants revealed that only one item yielded a significant between-group difference.

With regard to trust, speech-based interaction was associated with significantly higher general trust in technological systems (Q1: $p = 0.019$, $d = 0.59$). This suggests that auditory interaction may activate a broader sense of confidence in the technology, possibly due to the naturalness and immediacy of spoken dialogue. However, no significant differences were found for perceived truthfulness (Q8: $p = 1.000$) or expressed doubt about system accuracy (Q7: $p = 0.082$), indicating that modality does not influence situation-specific credibility assessments. Developers should therefore not assume that adding a voice interface will make users perceive AI-generated content as more truthful or accurate.

In terms of usefulness, no significant differences were observed across any item. Both modalities equally supported knowledge acquisition (Q2: $p = 1.000$) and willingness to reuse the system (Q9: $p = 0.515$). Directional trends, though not statistically significant, suggested that text-based output may be associated with higher intentions to share information (Q3: $p = 0.131$) and apply it in academic contexts (Q4: $p = 0.066$). These tendencies are consistent with the practical advantage of written text for retention and reuse, but they require replication in larger, more diverse samples before practical recommendations can be derived.

For ease of use, no significant differences emerged between the two conditions. Participants evaluated both interaction modes as equally accessible, demonstrating that speech-based interfaces do not introduce additional usability barriers despite their lower familiarity for many users.

Several limitations should be acknowledged when interpreting these results. The sample comprised $N = 66$ participants, predominantly students and employees of a German university of applied sciences (mean age 23.51 years). This sample is relatively homogeneous with respect to age, educational background, and cultural context. Age and prior experience with AI systems are known to moderate technology acceptance and trust formation [4], and cultural dimensions, such as individualism versus collectivism or uncertainty avoidance, may substantially affect how users respond to voice-based AI interfaces. Future research should replicate this study with larger, more diverse, and cross-cultural samples to improve the generalizability of the findings.

A further limitation concerns the comparability of the two experimental conditions. The voice-based prototype was implemented using a custom integration of Microsoft Azure Text-to-Speech (TTS) and Speech-to-Text (STT) services with the ChatGPT API, operated via a C++ application. This differs substantially from the standard ChatGPT web interface used in the text-based condition. Consequently, observed differences between conditions may partially reflect factors beyond interaction modality alone, such as differences in response latency, voice naturalness, or interface familiarity. Future research should employ more equivalent technical implementations, for instance by leveraging ChatGPT's native voice mode, to isolate modality effects more cleanly. Nonetheless, this limitation is acknowledged transparently here, as it motivates important directions for methodological refinement in subsequent studies.

REFERENCES

[1] OECD, Ed., Artificial Intelligence in Society. Paris: OECD Publishing, 2019. doi: 10.1787/eedfee77-en.

[2] T. B. Brown et al., "Language Models are Few-Shot Learners," 2020, arXiv. doi: 10.48550/ARXIV.2005.14165.

[3] K. A. Hoff and M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," Hum.

Factors, vol. 57, no. 3, pp. 407–434, May 2015, doi: 10.1177/0018720814547570.

[4] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis, "User Acceptance of Information Technology: Toward A Unified View," *MIS Quarterly*, vol. 27, no. 3, pp. 425–478, Sep. 2003, doi: 10.2307/30036540.

[5] R. C. Mayer, J. H. Davis, and F. D. Schoorman, "An Integrative Model of Organizational Trust," The Academy of Management Review, vol. 20, no. 3, p. 709, Jul. 1995, doi: 10.2307/258792.

[6] C. I. Nass and S. Brave, Wired for speech: how voice activates and advances the human-computer relationship. in Computer-human interaction. Cambridge, Mass.: MIT Press, 2005.

[7] T. Bickmore and T. Giorgino, "Health dialog systems for patients and consumers," Journal of Biomedical Informatics, vol. 39, no. 5, pp. 556–571, Oct. 2006, doi: 10.1016/j.jbi.2005.12.004.

[8] S. Oviatt, "Ten myths of multimodal interaction," Commun. ACM, vol. 42, no. 11, pp. 74–81, Nov. 1999, doi: 10.1145/319382.319398.

[9] A. Purington, J. G. Taft, S. Sannon, N. N. Bazarova, and S. H. Taylor, "'Alexa is my new BFF': Social Roles, User Satisfaction, and Personification of the Amazon Echo," in Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver Colorado USA: ACM, May 2017, pp. 2853–2859. doi: 10.1145/3027063.3053246.

[10] E. Adamopoulou and L. Moussiades, "An Overview of Chatbot Technology," in Artificial Intelligence Applications and Innovations, vol. 584, I. Maglogiannis, L. Iliadis, and E. Pimenidis, Eds., in IFIP Advances in Information and Communication Technology, vol. 584, Cham: Springer International Publishing, 2020, pp. 373–383. doi: 10.1007/978-3-030-49186-4_31.

[11] T. Ueno, Y. Sawa, Y. Kim, J. Urakami, H. Oura, and K. Seaborn, "Trust in Human-AI Interaction: Scoping Out Models, Measures, and Methods," in CHI Conference on Human Factors in Computing Systems Extended Abstracts, New Orleans LA USA: ACM, Apr. 2022, pp. 1–7. doi: 10.1145/3491101.3519772.

[12] L. Marconi, L. Longo, and F. Cabitza, "Assessing Interaction Quality in Human–AI Dialogue: An Integrative Review and Multi-Layer Framework for Conversational Agents," MAKE, vol. 8, no. 2, p. 28, Jan. 2026, doi: 10.3390/make8020028.

[13] S. W. T. Ng and R. Zhang, "Trust in AI chatbots: A systematic review," Telematics and Informatics, vol. 97, p. 102240, Feb. 2025, doi: 10.1016/j.tele.2025.102240.

[14] C. Heckersbruch, A. Öksüz, N. Walter, J. Becker, and G. Hertel, "Vertrauen und Risiko in einer digitalen Welt," 2013.

[15] Y. Ye, H. You, and J. Du, "Improved Trust in Human–Robot Collaboration with ChatGPT," 2023, arXiv. doi: 10.48550/ARXIV.2304.12529.

[16] E. M. Rogers, Diffusion of innovations, Fifth edition. New York London Toronto Sydney: Free Press, 2003.

[17] D. Huyler and C. M. McGill, "Book Review: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, by CreswellJohn and CreswellJ. David. Thousand Oaks, CA: Sage Publication, Inc.275 pages, $67.00 (Paperback).," New Horizons in Adult Education and Human Resource Development, vol. 31, no. 3, pp. 75–77, Jun. 2019, doi: 10.1002/nha3.20258.

[18] A. Onwuegbuzie, N. Leech, and K. Collins, "Qualitative Analysis Techniques for the Review of the Literature," TQR, Jan. 2015, doi: 10.46743/2160-3715/2012.1754.

[19] E. Brynjolfsson and T. Mitchell, "What can machine learning do? Workforce implications," *Science*, vol. 358, no. 6370, pp. 1530–1534, Dec. 2017, doi: 10.1126/science.aap8062.

[20] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," MIS Quarterly, vol. 13, no. 3, p. 319, Sep. 1989, doi: 10.2307/249008.

[21] L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," Psychometrika, vol. 16, no. 3, pp. 297–334, Sep. 1951, doi: 10.1007/BF02310555.

[22] A. Gupta, D. Basu, R. Ghantasala, S. Qiu, and U. Gadiraju, "To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System," in Proc. ACM Web Conference 2022 (WWW '22), Lyon, France: ACM, Apr. 2022, pp. 1–10. doi: 10.1145/3485447.3512248.

[23] M. Rheu, J. Y. Shin, W. Peng, and J. Huh-Yoo, "Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design," Int. J. Hum.–Comput. Interact., vol. 37, no. 1, pp. 81–96, 2021. doi: 10.1080/10447318.2020.1807710.