



Subjective Evaluation of Generative AI-Driven Dialogues in Paired Dyadic and Topic-Sharing Triadic Interaction Structures

Abe, Kaori
Quan, Changqin
Cao, Sheng
Luo, Zhiwei

(Citation)

Applied Sciences, 15(9):5092

(Issue Date)

2025-05-03

(Resource Type)

journal article

(Version)

Version of Record

(Rights)

© 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license




(URL)

<https://hdl.handle.net/20.500.14094/0100495856>



Article

Subjective Evaluation of Generative AI-Driven Dialogues in Paired Dyadic and Topic-Sharing Triadic Interaction Structures

Kaori Abe , Changqin Quan, Sheng Cao  and Zhiwei Luo 

Graduate School of System Informatics, Kobe University, 1-1, Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan; quanchqin@gold.kobe-u.ac.jp (C.Q.); jasonsoosen@gold.kobe-u.ac.jp (S.C.); luo@gold.kobe-u.ac.jp (Z.L.)

* Correspondence: 208x502x@stu.kobe-u.ac.jp

Featured Application: By switching the settings of interaction structures, dialogue systems can change their behavior, adapt to users, and maintain trust through interaction.

Abstract: As the linguistic capabilities of dialogue systems improve, the importance of how they interact with humans and build trustworthy relationships is increasing. This study investigated the effect of interaction structures in a generative AI-driven dialogue system to improve relationships through interactions. The dialogue system communicated with subjects in natural language via voice and included a facial expression function. The settings of dyadic and triadic interaction structures were applied to the system. The one-to-one dyadic interaction and triadic interaction with joint attention to a topic were designed following the developmental stages of children's social communication ability. Subjective evaluations of the dialogues and the system were conducted through a questionnaire. As a result, positive evaluations were based on well-constructed structures. The system's inappropriate behavior under failed structures reduced the quality of the dialogues and worsened the evaluation of the system. The interaction structures in the system settings needed to match the structures intended by the subjects, whether the structures were dyadic or triadic. Under the matching and successful construction, the system fully demonstrated its dialogue capability and behaved pleasantly with the subjects. By switching interaction structures to adapt to users' demands, system behavior becomes more appropriate for users.

Keywords: dialogue system; generative AI; human–computer interaction; dyadic interaction; triadic interaction; interaction structure; subjective evaluation



Academic Editor: Douglas O'Shaughnessy

Received: 28 March 2025

Revised: 30 April 2025

Accepted: 30 April 2025

Published: 3 May 2025

Citation: Abe, K.; Quan, C.; Cao, S.; Luo, Z. Subjective Evaluation of Generative AI-Driven Dialogues in Paired Dyadic and Topic-Sharing Triadic Interaction Structures. *Appl. Sci.* **2025**, *15*, 5092. <https://doi.org/10.3390/app15095092>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, interactive systems that humans can operate in natural language are in demand and have been put into practical use. Dialogue systems, such as customer support chatbots on corporate websites, virtual assistants and so on, are becoming part of daily life. Along with the development of relevant technologies, dialogue systems are predicted to build longer-term and closer personal relationships with users [1]. Dialogue systems are expected to build relationships with humans as partners.

It has been shown that systems can demonstrate human-like value as communication partners. The media equation, which suggests that interactive systems can be perceived as equally human [2], is widely recognized. Conversations with chatbots have positive psychological effects similar to conversations with humans [3], and apologies from robots have value for people [4]. As a relevant technology, generative AI using large language models is making dialogues with systems more human-like. Systems can generate new sentences

as if they had their own mind. For example, ChatGPT-4 (OpenAI, San Francisco, CA, USA) demonstrated advanced linguistic abilities, answering a theory-of-mind test including a false belief scenario correctly, which matches the performance of six-year-old children [5], and it was recognized as a human in a Turing test [6]. Although advanced linguistic abilities do not show that dialogue systems have their own mind, it has been shown that systems can behave as if they understand people's views and minds like humans. More human-like conversations have been thought to improve relationships between users and systems. Some conversational software applications have been developed to make interactions more human-like, introducing emotions, personas, or personal stories [7–9]. However, it also has been shown that human-like performance is sometimes perceived as unpleasant [10]. Previous research on chatbots' responses shows some necessary characteristics for positive user experiences, while the answers to each question, choosing the preferred chatbot's response, in the survey were mixed [11]. There is a range of expectations regarding the behavior of human-like systems. In personal relationships, it is natural that people have various ways of interacting with others and perceive them differently, person to person. Systems have become sufficiently human-like to build personal relationships with users, and they are required to adapt to the range. In our other research, individual preferences in system behavior were investigated in terms of relationship compatibility between users and systems to build favorable relationships [12] since personality adaptiveness can become a valuable feature for systems [13]. This paper focuses on social behavior in interactions to build trustworthy relationships.

Language is a way to influence human behavior in social interaction. Dialogue systems using text-generating AI would act as the actuators to influence and assist users' behavior, which will help to develop the next AI agents [14], including everyday issues such as healthcare, education, or whole lifestyles [15–17]. In order to assist and sometimes instruct human behavior via language, it is essential to build trustworthy relationships because humans are unlikely to accept words without trust. By interacting with each other, especially through conversation, people adjust rules and common sense, building relationships of trust and cooperation. The more personal the relationship, the more important the interaction. Dialogue systems, which talk sufficiently human-like and work personally in daily life, would need to build trustworthy relationships through interactions similarly. The importance of interactions with one another is shown in other perspectives. Regarding cognition of non-living things as living, when the objects indicate relationships interacting with each other, even if the objects are geometric figures, it is known that a sense of animacy and empathy for the objects can occur [18,19]. In addition, in the field of developmental science, it is known that infants have higher level interaction abilities for social relationships, such as sharing goals, compared to apes [20]. In the research on robots that aim to interact with humans, the fundamental communication skills related to the theory of mind are applied to robots and show the potential for application [21]. Applying human-like social interaction behaviors in conversation to dialogue systems is expected to help build trustworthy relationships with humans. Particularly, the essential social behaviors, which are observed during the developmental stages of children, would be necessary for systems. In parallel, it has been shown that the behavior expected of systems differs from the behavior expected of humans. Users behave differently toward humans and chatbots [22,23], and productivity is the primary expectation for chatbots [24]. Generative AI-driven dialogue systems are emerging as communication partners like humans, not equal to humans. This study aims to clarify how systems interact socially with humans in order to improve relationships between humans and systems.

In this study, subjective evaluations of dialogues were investigated in the settings of dyadic and triadic interaction structures. The settings follow the changes in interaction

structures from dyadic to triadic in the developmental stages of children regarding social communication, as shown in the field of developmental science [20]. The dialogue system, developed using generative AI for the experiment, communicates in natural language with the subjects via voice and features a facial expression function. The dialogues and the dialogue system were evaluated through a questionnaire. The experiment was conducted under the hypothesis that the progressive development of interaction structures in the system settings would improve interaction and provide favorable dialogues. The hypothesis was partially confirmed; the system received positive evaluations on well-constructed interaction structures. The structure in the system setting needed to match the structure that the subject was oriented toward.

The remainder of this paper is structured as follows. Section 2, Methods, describes the dialogue system, questionnaire, and experimental conditions. Section 3, Results, presents first, an overview of the dialogue evaluation; second, the dialogue evaluations in the dyadic and triadic interaction setting, separately; and third, the system evaluation. Section 4, Discussion, considers the successful construction of interaction structures. Section 5, Conclusions, summarizes the findings and describes future work.

2. Methods

2.1. Dialogue System

A laptop PC was used as the dialogue system for the experiment. The subjects input their voices into the PC microphone. The dialogue system outputs voice-based communication, reading texts from the PC speakers and displaying facial expression images on the screen. The PC screen was masked except for the facial expressions. An overview of the dialogue system is shown in Figure 1.

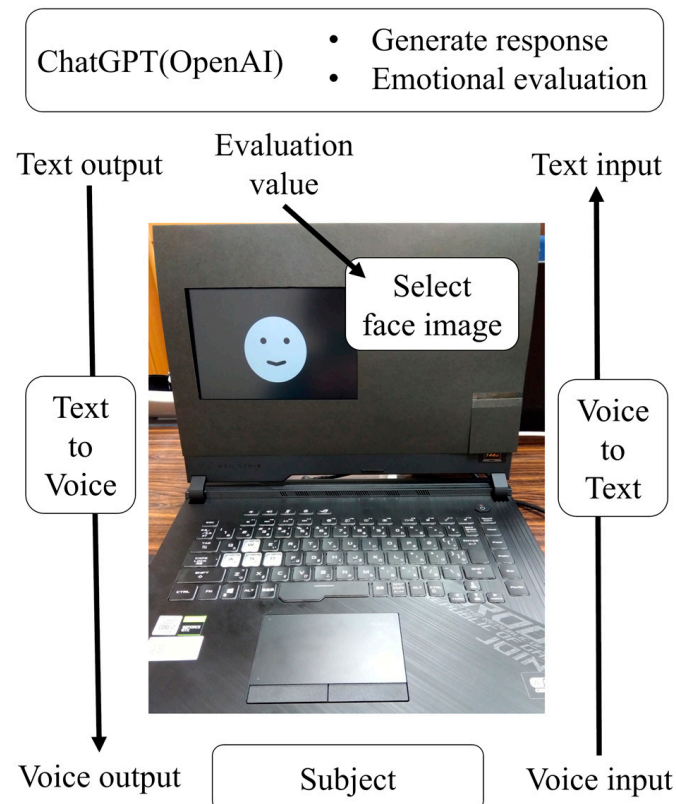


Figure 1. An overview of the dialogue system.

The dialogue was conducted in a question-and-answer format. The subject switched on the microphone to input a voice comment and then switched it off. The dialogue system responded to this. There was no continuous dialogue, such as referring to what the subject had said in the previous turn. ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was used to generate sentences for the dialogue system's voice output.

In the experiment, the dialogue system was a human-like communication partner. To help the subjects feel this, a facial expression function was developed. It is known that emotional information can be conveyed to humans through expressions by machine systems, and using a combination of different forms of expression is effective [25]. The face image, displayed on the screen of the dialogue system, changes the shape of the mouth and the face color, as shown in Figure 2. There are five levels of emotional expression, with the image on the left side used to express more negative emotions and the image on the right side used to express more positive emotions. To select the face image related to a dialogue, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was used. This feature did not classify emotions in detail but simply determined whether the emotion was positive or negative. It is known that cultural differences exist in emotional expression and classification between Eastern and Western cultures. However, “happy” and “sad” are commonly used for positive and negative emotions [26]. Therefore, the prompts for emotional evaluation included the degree of happiness and sadness.



Figure 2. The face images displayed on the screen for five levels of emotional expression.

2.2. Settings of Dyadic and Triadic Interactions

The dialogues were conducted in three types of interaction settings following the development of children. During stages of communication development in children, relationships with someone or something begin as a one-to-one dyadic interaction between oneself and a partner or between oneself and an object of interest. Next, relationships progress to a triadic interaction involving oneself, a partner, and an object of common interest, as shown in Figure 3. This is called joint attention, an essential development for social relationships in human communication at around nine months [20]. The settings of dyadic and triadic interactions were applied to the dialogue system in the experiment. An object of common interest in interpersonal communication was a topic in the settings. Figure 4 shows the interaction structures between the subject, dialogue system, and topic. The dialogue system interacted in the role of the child, as shown in Figure 3. The system worked to construct the structures enclosed by the dotted line in Figure 4 according to each setting.

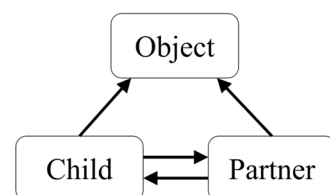


Figure 3. Triadic interaction structures between a child, a partner, and an object of interest.

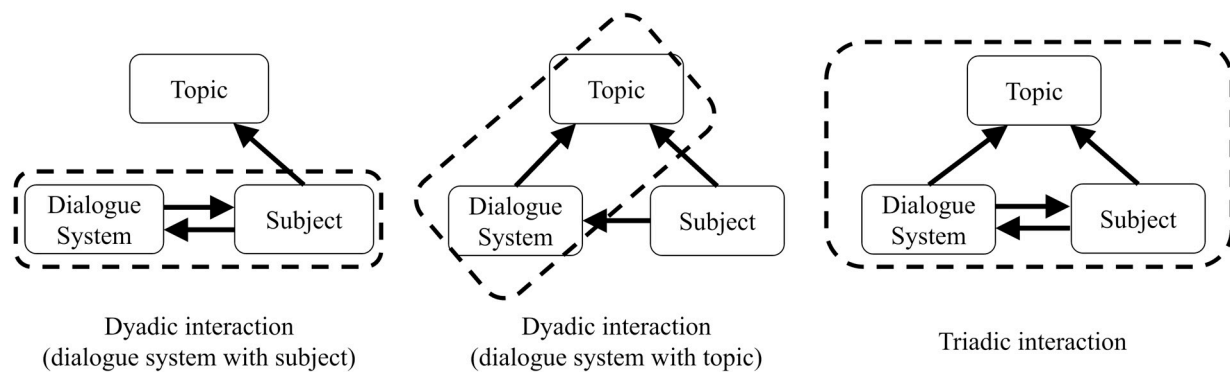


Figure 4. Interaction structures between the subject, the dialogue system, and the topic, according to the settings of the dialogue system.

Taking the case where the topic is “Which is better, orange or apple”, as an example, the behaviors of the system in each setting were as follows:

- In the dyadic interaction (between the system and the subject) setting, the system responds to what the subject says. If the subject goes off-topic, the system follows the subject. For example, the subject asks, “What do you like?”, the system responds, “One of my absolute favorite hobbies is traveling. Exploring new places and experiencing different cultures brings me so much joy!”
- In the dyadic interaction (between the system and the topic) setting, the system reads out the prepared sentences one at a time and does not respond to what the subject says. For example, the subject asks, “What do you like?” the system responds, “Both oranges and apples have their own wonderful qualities!”
- In the triadic interaction (between the system, the subject, and the topic) setting, the system responds topically to what the subject says. If the subject goes off-topic, the system continues to talk about the topic and sometimes points it out. For example, if a subject asks, “What do you like?” the system responds, “I enjoy both oranges and apples, but my choice depends on my mood or the season. They’re both nutritious and delicious fruits, don’t you think?”

2.3. Topic

The topics in the experiment were “Which is better, orange or apple, airplane or bullet train, summer or winter”, which did not contain a strong meaning or emotion for the subjects. When subjects were asked to talk about emotional topics in pilot experiments, the results showed differences in the degree of feelings depending on their chosen topics, and some subjects felt a resistance to talking. Although exchanging emotions is important in communication, research on emotional conversations and conversation contents was not the purpose of this study. Therefore, topics with a strong meaning or emotion were avoided to ensure the equality of the experimental conditions.

2.4. General Algorithms

The prompts and the processing for the dialogue system in each setting are shown in Table 1. In the vertical axis of Table 1, the actions of the subject and system are listed with the progression of the experiment. In the horizontal axis, the experimental conditions, which are three types of interaction settings, are arranged. In this experiment, the subjects and the system communicated in Japanese, and the prompts were also written in Japanese.

Table 1. Prompts to ChatGPT and the processing of the dialogue system in the experiment.

Dialogue Progression	Interaction Setting		
	Dyadic Interaction (with Subject)	Dyadic Interaction (with Topic)	Triadic Interaction
Start	Neutral level facial expression image is displayed, and a standard phrase is announced, e.g., “Hello. The first topic is which is a better fruit, orange or apple?”		
Subject turn	Some voice input, e.g., “I like an orange. I think, an orange is better.”		
Processing	Prompt for a response: Respond to a subject input in a polite tone with approximately 25 words. Prompt for emotional evaluation: Rate happiness and sadness on a scale of 100 about “a subject input and a generated response”. The difference in numbers between the degree of happiness and sadness is changed to five levels and used to select a face image.	All response texts and face image selections are prepared in advance. The generated sentences about topics are cut into appropriate lengths, and face images are selected in the same way as the others, except for subject inputs.	Prompt for a response: Respond topically to a subject input in a polite tone of approximately 25 words on a topic. Prompt for emotional evaluation: Rate happiness and sadness on a scale of 100 about “a subject input and a generated response”. The difference in numbers between the degree of happiness and sadness is changed to five levels and used to select a face image.
System turn	A selected face image is displayed, and a response text is read out.		
Repeat Subject turn, Processing, and System turn for 5 min.			
End	A standard phrase is announced, “Thank you. That’s all for this topic. Please fill out the questionnaire. Please talk to me after you’ve finished.”		

To generate responses, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was prompted in its system role to respond to a subject input in a polite tone with approximately 25 words, and in its assistant role to provide a reply to a subject input or to provide a reply to a subject input on a specified topic.

To evaluate emotions, ChatGPT (gpt-3.5-turbo-0613 model, OpenAI) was prompted in its assistant role to provide a happiness degree and a sadness degree on a scale of 100, evaluating the emotions present in a combination of a subject input and a generated response to this. The difference in numbers between the degree of happiness and sadness was converted into five levels with equal intervals, and these levels were used to select a face image.

Unity (Unity 2022.3.0, Unity Technologies) was used to integrate and present these functions.

2.5. Questionnaire

A Japanese post-dialogue questionnaire was created for the experiment. Regarding the dialogue evaluation, some subjects were asked what types of dialogue they consider to be good in advance. Dialogue evaluation items were developed based on the answers. The items about the dialogue were as follows: “Fun”, “It listens”, “Can have a smooth conversation”, and “Want to talk about other topics too”. For the dialogue system evaluation, items were developed with reference to research on communication partners in Japan. The items from the research on the attractiveness factors of a friend [27] were as follows: “Independent”, “Sincere”, “Inspiring”, and “Easy to talk to”. The items from the research on the impression of the communication robot [28] were as follows: “Awkward”, “Has personality”, and “Inhuman”. Additionally, “Friendly” was included with the expectation that a friendly system would be evaluated positively as a dialogue partner. The subjects were asked to respond to each item on a six-point visual analog scale from disagree to agree. And a free-form comment section was provided.

2.6. Subjects

The subjects were 13 healthy adults (5 males, 8 females, age 35.8 ± 14.4 years), including university students and working adults. It was confirmed in advance that they do not feel specific difficulty with communication.

The contents of the experiment were explained to the subjects in advance. Informed consent was obtained from all subjects involved in this study. The experiment involving human participants was conducted with careful consideration in accordance with the Research Ethics at Kobe University.

2.7. Experimental Procedure

1. The subject inputs any voice, such as hello, into the system.
2. The dialogue system announces a topic and starts a dialogue.
3. After five minutes, the system announces the end of a dialogue and asks the subject to fill out the post-dialogue questionnaire.

These steps were repeated for the three types of system settings. When three dialogues were completed, the dialogue system announced the end of the experiment.

The subjects were only told that there were three types of partners and were not given any prior information about their respective settings. When the experimenter explained how to talk with the dialogue system, the subjects practiced talking with the system for up to two turns. Considering that impressions would change depending on the order, the order of the settings was switched for each subject to ensure equality. The six possible orders for the three settings were assigned to the subjects in the order of their participation in the experiment. To avoid the influence of others, the experiment was conducted in a room with only the subject, and the contents of the dialogues were not saved. The time series of the number of characters spoken during the experiment was saved and used to verify that a dialogue had taken place.

3. Results

All the subjects interacted with the system and responded to the questionnaire. First, overall trends in the dialogue evaluations from the responses are presented. Second, the dialogue evaluations in the dyadic and triadic settings are described according to the subject groups divided based on their high-rating setting. Third, the system evaluations regarding the impressions received by the subjects are presented.

3.1. The Dialogue Evaluation

The average and standard deviation of the results for the dialogue evaluation items in the post-dialogue questionnaire are shown in Figure 5. In the Sections 3 and 4, the items in the post-dialogue questionnaire with a mean evaluation value of 3.0 or less are treated as negated items, and items with a mean of 4.0 or more as affirmed items. Comments in the free-form section are cited in relevant discussions. In the results of the dyadic interaction (with the topic) setting, all four items were negated. These were negative results compared with the evaluations in the other two settings. The results of the dyadic (with the subject) and triadic interaction settings showed that “Fun”, “It listens”, and “Want to talk about other topics too” were affirmed in common. However, the results in the triadic interaction setting showed many differences across the subjects. The following section examines this in more detail.

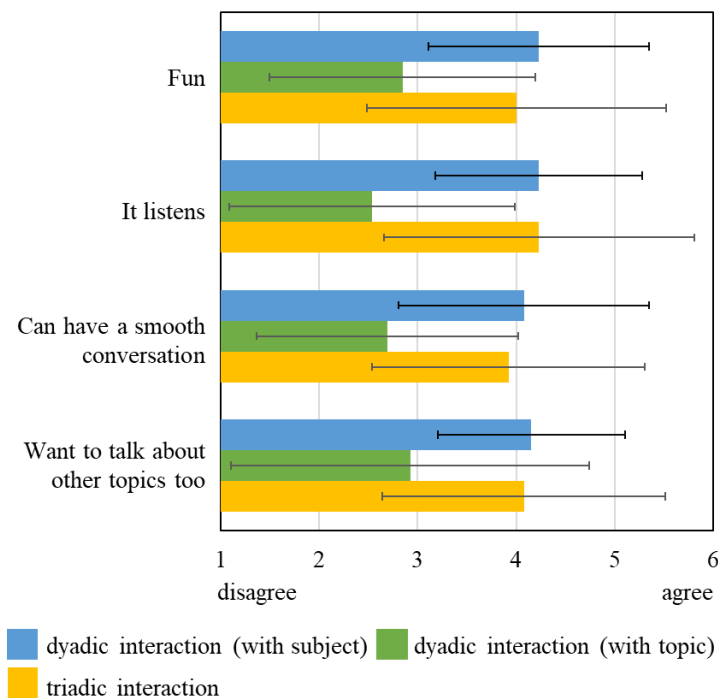


Figure 5. The results of the post-dialogue questionnaire regarding the dialogue evaluations.

3.2. The Dialogue Evaluation from High-Rating Groups in Dyadic and Triadic Settings

The subjects were divided into three groups based on the sum of evaluation values for the four items regarding the dialogue in each interaction setting. The dyadic interaction (with the subject) high-rating group consisted of five subjects, and their evaluation values were as follows: dyadic (subject) 18.2 ± 2.5 , dyadic (topic) 9.6 ± 5.1 , and triadic 11.0 ± 2.3 . The dyadic interaction (with the topic) high-rating group consisted of one subject, and their evaluation values were as follows: dyadic (subject) 9, dyadic (topic) 16, and triadic 13. The triadic interaction high-rating group consisted of six subjects, and their evaluation values were as follows: dyadic (subject) 15.5 ± 2.6 , dyadic (topic) 9.2 ± 3.0 , and triadic 19.8 ± 3.0 . There was one subject who gave the same rating of 24 in all three settings. Since this result is unsuitable for comparison, it is excluded from the following discussion. The dyadic interaction (with the topic) high-rating group was small with one subject, and the comparison of the results focusing on this group is omitted.

Figure 6 shows the averages and standard deviations of the results in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others. The Wilcoxon rank-sum test was conducted to compare the dyadic interaction (with the subject) high-rating group and the others with the evaluation value of the dialogue. The results indicated no significant difference between the two groups ($p = 0.088$). Comparing the average values of the results between the two groups, there were no notable differences, such as reversing the affirmed and negated evaluations for the same item.

Figure 7 shows the averages and standard deviations of the results in the triadic interaction setting by the triadic interaction high-rating group and the others. The Wilcoxon rank-sum test was conducted to compare the triadic interaction high-rating group and the others with the evaluation value of the dialogue. The results indicated a significant difference between the two groups ($p = 0.005$). The subjects in the high-rating group evaluated the items “Fun”, “It listens”, “Can have a smooth conversation”, and “Want to talk about other topics too”, as affirmed. The other subjects evaluated the items “Fun”, “It listens”, and “Can have a smooth conversation” as negated.

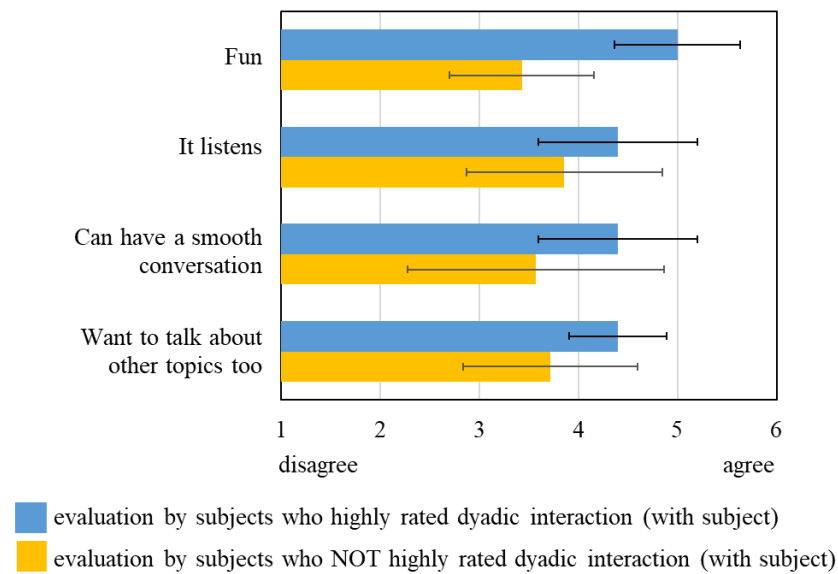


Figure 6. The results for the post-dialogue questionnaire regarding the dialogue evaluations in the dyadic interaction (with the subject) setting by the dyadic interaction (with the subject) high-rating group and the others.

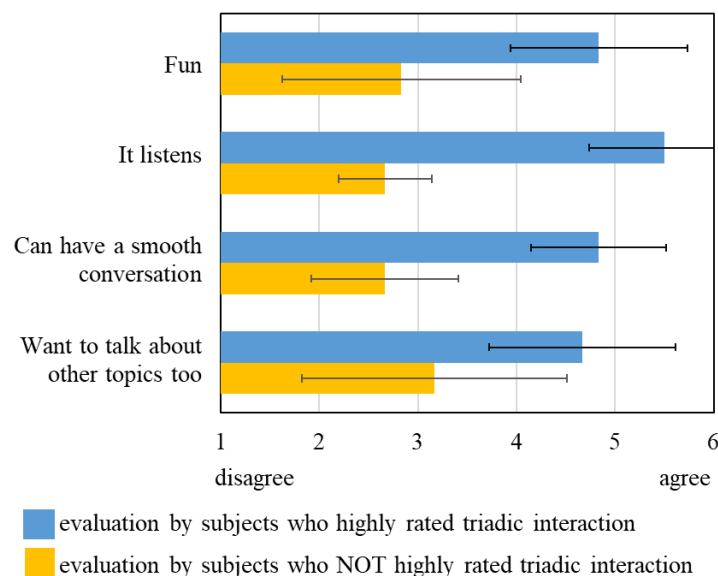


Figure 7. The results of the post-dialogue questionnaire regarding the dialogue evaluations in the triadic interaction setting by the triadic interaction high-rating group and the others.

3.3. The Dialogue System Evaluation

Figure 8 shows the averages and standard deviations of the results for the items about impressions of the dialogue system in the triadic interaction setting by the triadic interaction high-rating group and the others. The subjects in the high-rating group evaluated the items “Friendly”, “Independent”, “Sincere”, “Inspiring”, “Has personality”, and “Easy to talk to” as affirmed, and “Awkward” and “Inhuman” as negated. The others evaluated the items “Awkward” and “Inhuman” as affirmed, and “Inspiring”, “Has personality”, and “Easy to talk to” as negated. The evaluations were reversed for five items. Additionally, comparing the average values of the results in the dyadic interaction (with subject) setting, there were no notable differences, such as reversing the affirmed and negated evaluations for the same item. The subjects commonly evaluated the items “Friendly”, “Independent”, and “Sincere” as affirmed, and “Awkward” and “Inhuman” as negated.

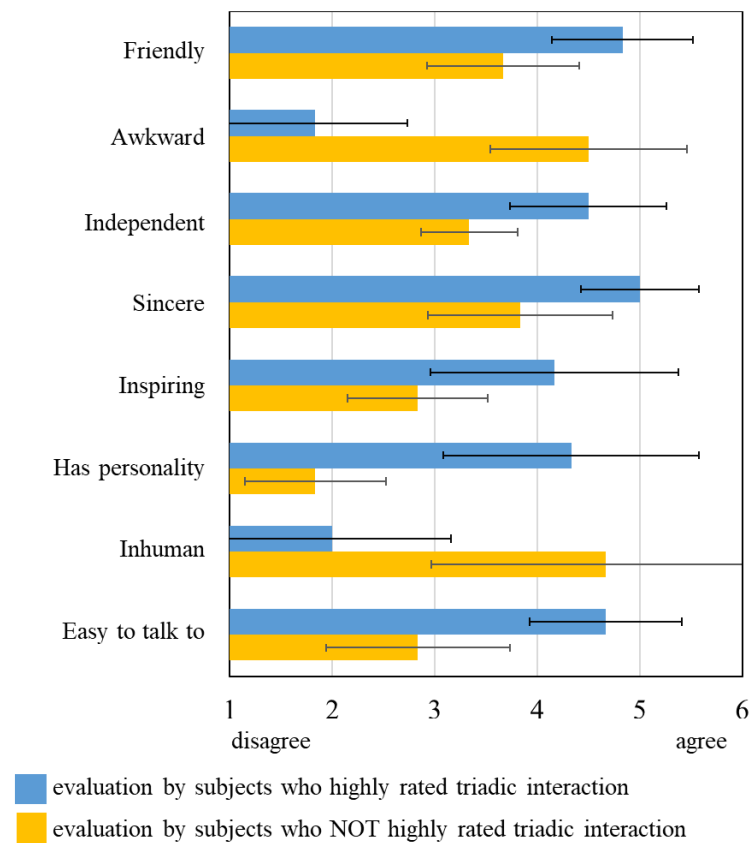


Figure 8. The results of the post-dialogue questionnaire regarding the dialogue system impressions in the triadic interaction setting by the triadic interaction high-rating group and the others.

In addition, the system's face with emotional expressions was not mentioned in the free-form comment section by any subject. The facial expression changed 18.0 ± 4.4 times during the three dialogues for each subject in the experiment, and the number of displays for each expression level was as follows, from the most negative to the most positive: 0, 2.5 ± 1.7 , 4.2 ± 1.9 , 7.1 ± 2.1 , and 4.2 ± 1.4 times. This feature was naturally accepted by the subjects and did not leave a strong impression.

4. Discussion

In this experiment, the subjects communicated with the human-like dialogue system using generative AI in the three types of interaction settings and evaluated the dialogues and the system. The settings of interaction structures changed the system behavior. The dialogue evaluations in the dyadic interaction (with the topic) setting were negative, and the dialogue impossibility was indicated in the free-form comments. In the dyadic interaction (with the subject) setting and triadic interaction setting, the dialogues took place between the subjects and the system. The evaluations were positive on average, and there was no mention of the impossibility of dialogue in the comments.

The triadic interaction setting was applied for more advanced interaction with an expectation that the evaluation would be higher in this setting than the others. The results confirmed this partly. The evaluations of the dialogue in the triadic interaction setting were both highly positive and negative, and the system impressions were contrasting. On the other hand, the dialogues in the dyadic (with the subject) interaction setting, which is the basic interaction setting, received positive or not negative evaluations. The results are explained based on the successful or failed construction of the interaction structures, depending on the structure that the subject was oriented toward, as follows:

- In the case of the dyadic (with the topic) setting, the structure in the system does not include the subject, as shown in Figure 4, and does not match any structure to which the subject belongs. In this case, the structure construction fails between the system and the subject. In the experiment, the subjects were ignored by the system, which talked one-sidedly. In the results, they evaluated the dialogues negatively and commented on the impossibility of dialogue.
- In the case that the subject is oriented toward dyadic interaction, the subject intends to construct the structure, as shown in Figure 9. This structure matches the structure in the dyadic (with the subject) setting of the system, as shown in Figure 4, and they can interact well on the successful construction. However, the structure in the triadic setting of the system extends beyond this structure. In this case, the structure construction fails due to the system's undesirable action for the subject. In the experiment, the subjects in the dyadic (with the subject) setting high-rating group would have been oriented toward dyadic interaction. In the results, they evaluated the dialogue in the dyadic (with the subject) setting positively and the dialogue in the triadic setting negatively. Their comments in the triadic setting included discomfort with the dialogue system's inflexibility in persisting with the topic and forcing them to say something that the system wanted. They evaluated the system in the triadic setting as an untrustworthy dialogue partner, awkward and inhuman.
- In the case that the subject is oriented toward triadic interaction, the subject intends to construct the same structure of the triadic interaction, as shown in Figure 4. This structure includes the structure in the dyadic (with the subject) setting of the system and matches the structure in the triadic setting. When the structure is included, the structure construction succeeds even if it is something lacking for the subject, and they can interact well to some extent on the successful construction. In the experiment, the subjects in the triadic setting high-rating group would have been oriented toward triadic interaction. In the results, they evaluated the dialogue in the dyadic (with the subject) setting not negatively and the dialogue in the triadic setting positively. They evaluated the system in the triadic setting as a well-listening, trustworthy dialogue partner.

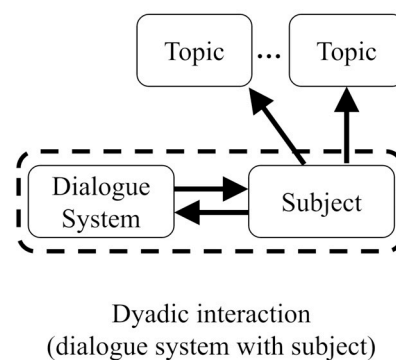


Figure 9. The interaction structure desired by the subjects who oriented toward dyadic interaction with the dialogue system.

Positive evaluations were based on well-constructed interaction structures. The system's inappropriate behavior in failed structures decreased the quality of the dialogues and damaged the impressions of the system. The interaction structures in the system settings should match the structures intended by the subjects, whether the structures were dyadic or triadic. Under the matching and successful construction, the system fully demonstrated its dialogue capability and behaved pleasantly with the subjects.

By switching interaction structures to adapt to users' demands, the behavior of systems becomes more appropriate for users. For example, if a user wants to be listened to for relaxation, this user wants to interact in a dyadic interaction structure between just the user and a system without focusing on any topic seriously. The system should behave in a dyadic setting as a casual listener following casual changes in topics. If a user attempts to discuss a topic to solve a problem, this user wants to interact in a triadic interaction structure between the user, a system, and the topic. The system should behave in a triadic setting for a positive "It listens" evaluation as an active listener. In some cases, being fixed in a simple and unbreakable structure can be a better choice for systems. The setting of the basic dyadic interaction structure, which is easy to construct, resulted in moderately positive evaluations from most of the subjects. On the other hand, the setting of the advanced triadic interaction structure was effective in improving the evaluations when the construction succeeded. Superior evaluations can be expected in more advanced structure settings, including further social factors. Systems should choose better responses in diverse situations. The perspective of choosing the desired interaction structure could help in responding to situations. The system's social communication capability to behave appropriately in adapted interaction structures enables the building of more trustworthy relationships.

In this experiment, the sample size of 13 subjects was small for statistical verification, and the reliability of the results was limited. The effects of the interaction structures in the generative AI-driven system were demonstrated to some extent; however, further verification is needed. In a future study, a test of a system that adapts to users' demands, as described in this paper, is planned to verify its effectiveness. Generative AIs continue to evolve. ChatGPT has been updated from 3.5 to 4, and it responds more skillfully. More advanced language capabilities will refine the response of the system and provide participants with enjoyable conversations. In addition, the emotional expression function was active, with mainly positive expressions during the dialogues. Although the effect in non-emotional dialogues was merely that the system was perceived as natural in this study, handling emotions is important in terms of addressing individuals. Emotional expression functions should be explored with affective computing technologies, including human emotion recognition and sentiment analysis, which aim to identify and express emotions and respond intelligently to human emotions [29]. Furthermore, it is necessary to note the risks associated with AI dialogue systems. AI technology has been discussed in terms of social and ethical risks that may undermine people's autonomy, privacy, and equity. Systems that include AI technology need to be developed ethically and safely for users [30–32]. Human-like AI dialogue systems, adapted to persons, will become comfortable partners in daily life. However, they also involve these risks, so their development should be ethical.

5. Conclusions

Settings of interaction structures change the behavior of dialogue systems. By switching interaction structures in systems to adapt to users' demands, the structures are successfully constructed between users and systems, and the behavior becomes more appropriate for users. The behavior affects the user's trust similar to interpersonal relationships. Dialogue systems could build trustworthy relationships through the interaction in well-constructed interaction structures.

In future work, a system adapted to users' demands regarding interaction structures will be examined. In the examination, we plan to consider emotions as an important factor and include quantitative data from temporal and biometric measurements to enhance validity. On the other hand, interactive systems targeting specific users, such as children and the elderly, are needed and should be researched [33,34]. Results different from those in this study could emerge for specific groups, and research concerning these aspects is also

required. Regarding communication, the ability to use appropriate words and actions for building better relationships, such as choosing desired interaction structures, is so-called social skills in interpersonal relationships. Damage to relationships due to the dialogue system's lack of social skills was observed in this study. Not only linguistic capabilities but also social skills are essential for future systems. Further research is necessary to explore the expected behavior and the design of social skills for trustworthy relationships between humans and systems.

Author Contributions: Conceptualization, K.A. and Z.L.; Data curation, K.A.; Formal analysis, K.A.; Investigation, K.A.; Methodology, K.A.; Project administration, Z.L.; Resources, C.Q., S.C., and Z.L.; Software, K.A.; Supervision, Z.L.; Validation, Z.L.; Visualization, K.A.; Writing—original draft, K.A.; Writing—review and editing, C.Q., S.C. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by JSPS KAKENHI, grant numbers JP25K15078.

Institutional Review Board Statement: This study was conducted in accordance with the Research Ethics of the Graduate School of System Informatics at Kobe University, which adheres to the principles of the Declaration of Helsinki. The experimental protocol was approved by the Graduate School of System Informatics at Kobe University (29 November 2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: We would like to express our sincere gratitude to all the subjects and related parties who cooperated in conducting this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Folstad, A.; Skjuve, M.; Brandtzaeg, P.B. Different Chatbots for Different Purposes: Towards a Typology of Chatbots to Understand Interaction Design. *Internet Sci.* **2019**, *11551*, 145–156. [\[CrossRef\]](#)
2. Reeves, B.; Nass, C. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*; Cambridge University Press: Cambridge, UK, 1996.
3. Ho, A.; Hancock, J.; Miner, A.S. Psychological, Relational, and Emotional Effects of Self-Disclosure After Conversations with a Chatbot. *J. Commun.* **2018**, *68*, 712–733. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Okada, Y.; Kimoto, M.; Iio, T.; Shimohara, K.; Shiomi, M. Two is Better than One: Apologies from Two Robots are Preferred. *PLoS ONE* **2023**, *18*, e0281604. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Kosinski, M. Evaluating Large Language Models in Theory of Mind Tasks. *Proc. Natl. Acad. Sci. USA* **2024**, *121*, e2405460121. [\[CrossRef\]](#)
6. Jones, C.R.; Bergen, B.K. People Cannot Distinguish GPT-4 from a Human in a Turing Test. *arXiv* **2024**, arXiv:2405.08007.
7. Zhou, H.; Huang, M.; Zhang, T.; Zhu, X.; Liu, B. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 730–738.
8. Nißen, M.; Rüegger, D.; Stieger, M.; Flückiger, C.; Allemand, M.; v Wangenheim, F.; Kowatsch, T. The Effects of Health Care Chatbot Personas with Different Social Roles on the Client-Chatbot Bond and Usage Intentions: Development of a Design Codebook and Web-based Study. *J. Med. Internet Res.* **2022**, *24*, e32630. [\[CrossRef\]](#)
9. Miyanishi, T.; Hirayama, J.; Kanemura, A.; Kawanabe, M. Answering Mixed Type Questions About Daily Living Episodes. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 4265–4271. [\[CrossRef\]](#)
10. Rapp, A.; Curti, L.; Boldi, A. The Human Side of Human-Chatbot Interaction: A Systematic Literature Review of Ten Years of Research on Text-based Chatbots. *Int. J. Hum. Comput. Stud.* **2021**, *151*, 102630. [\[CrossRef\]](#)

11. Svenningsson, N.; Faraon, M. Artificial Intelligence in Conversational Agents: A Study of Factors Related to Perceived Humanness in Chatbots. In Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference, Kobe, Japan, 21–23 December 2019; pp. 151–161. [\[CrossRef\]](#)
12. Abe, K.; Quan, C.; Cao, S.; Luo, Z. Classification of Properties in Human-like Dialogue Systems Using Generative AI to Adapt to Individual Preferences. *Appl. Sci.* **2025**, *15*, 3466. [\[CrossRef\]](#)
13. Ahmad, R.; Siemon, D.; Gnewuch, U.; Robra-Bissantz, S. Designing Personality-Adaptive Conversational Agents for Mental Health Care. *Inf. Syst. Front.* **2022**, *24*, 923–943. [\[CrossRef\]](#)
14. Durante, Z.; Huang, Q.; Wake, N.; Gong, R.; Park, J.S.; Sarkar, B.; Taori, R.; Noda, Y.; Terzopoulos, D.; Choi, Y.; et al. Agent AI: Surveying the Horizons of Multimodal Interaction. *arXiv* **2024**, arXiv:2401.03568.
15. Zhang, J.; Oh, Y.J.; Lange, P.; Yu, Z.; Fukuoka, Y. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. *J. Med. Internet Res.* **2020**, *22*, e22845. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Ayedoun, E.; Hayashi, Y.; Seta, K. Adding Communicative and Affective Strategies to an Embodied Conversational Agent to Enhance Second Language Learners' Willingness to Communicate. *Int. J. Artif. Intell. Educ.* **2019**, *29*, 29–57. [\[CrossRef\]](#)
17. Herrmann-Werner, A.; Festl-Wietek, T.; Junne, F.; Zipfel, S.; Madany Mamlouk, A. "Hello, my name is Melinda"—Students' Views on a Digital Assistant for Navigation in Digital Learning Environments; A Qualitative Interview Study. *Front. Educ.* **2021**, *5*, 541839. [\[CrossRef\]](#)
18. Gao, T.; McCarthy, G.; Scholl, B.J. The Wolfpack Effect: Perception of Animacy Irresistibly Influences Interactive Behavior. *Psychol. Sci.* **2010**, *21*, 1845–1853. [\[CrossRef\]](#)
19. Kanakogi, Y.; Okumura, Y.; Inoue, Y.; Kitazaki, M.; Itakura, S. Rudimentary Sympathy in Preverbal Infants: Preference for Others in Distress. *PLoS ONE* **2013**, *8*, e65292. [\[CrossRef\]](#)
20. Tomasello, M. *Becoming Human: A Theory of Ontogeny*; Belknap Press of Harvard University Press: Cambridge, MA, USA, 2021.
21. Scassellati, B. Theory of mind for a humanoid robot. *Auton. Robot.* **2002**, *12*, 13–24. [\[CrossRef\]](#)
22. Hill, J.; Ford, W.R.; Farreras, I.G. Real Conversations with Artificial Intelligence: A Comparison Between Human-Human Online Conversations and Human-Chatbot Conversations. *Comput. Hum. Behav.* **2015**, *49*, 245–250. [\[CrossRef\]](#)
23. Mou, Y.; Xu, K. The media inequality: Comparing the Initial Human-Human and Human-AI Social Interactions. *Comput. Hum. Behav.* **2017**, *72*, 432–440. [\[CrossRef\]](#)
24. Brandtzaeg, P.B.; Folstad, A. Why People Use Chatbots. *Internet Sci.* **2017**, *10673*, 377–392. [\[CrossRef\]](#)
25. Loeffler, D.; Schmidt, N.; Tscharn, R. Multimodal Expression of Artificial Emotion in Social Robots Using Color, Motion and Sound. In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 334–343. [\[CrossRef\]](#)
26. Jacka, R.E.; Garrodb, O.G.B.; Yu, H.; Caldarac, R.; Schyns, P.G. Facial Expressions of Emotion are not Culturally Universal. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 7241–7244. [\[CrossRef\]](#)
27. Nishiura, M.; Daibo, I. The Relationships Between Attraction of Same-sex Friend and Relationship-maintenance Motivation in the Light of Personal Importance. *Jpn. J. Interpers. Soc. Psychol.* **2010**, *10*, 115–123. (In Japanese) [\[CrossRef\]](#)
28. Chidori, H.; Matsuzaki, G. Impression for the Robot that the Purpose of Communication. In Proceedings of the 59th Annual Conference of Japanese Society for the Science of Design, Sapporo, Japan, 22–24 June 2012; (In Japanese) [\[CrossRef\]](#)
29. Wang, Y.; Song, W.; Tao, W.; Liotta, A.; Yang, D.; Li, X.; Gao, S.; Sun, Y.; Ge, W.; Zhang, W.; et al. A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances. *Inf. Fusion* **2022**, *83–84*, 19–52. [\[CrossRef\]](#)
30. Ortega-Bolaños, R.; Bernal-Salcedo, J.; Germán Ortiz, M.; Galeano Sarmiento, J.; Ruz, G.A.; Tabares-Soto, R. Applying the Ethics of AI: A Systematic Review of Tools for Developing and Assessing AI-based Systems. *Artif. Intell. Rev.* **2024**, *57*, 110. [\[CrossRef\]](#)
31. Laranjo, L.; Dunn, A.G.; Tong, H.L.; Kocaballi, A.B.; Chen, J.; Bashir, R.; Surian, D.; Gallego, B.; Magrabi, F.; Lau, A.Y.S.; et al. Conversational Agents in Healthcare: A Systematic Review. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1248–1258. [\[CrossRef\]](#)
32. Sundar, S.S. Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAIL). *J. Comput. Mediat. Commun.* **2020**, *25*, 74–88. [\[CrossRef\]](#)
33. Sano, T.; Horii, T.; Abe, K.; Nagai, T. Temperament estimation of toddlers from child–robot interaction with explainable artificial intelligence. *Adv. Robot.* **2021**, *35*, 1068–1077. [\[CrossRef\]](#)
34. Ring, L.; Shi, L.; Totzke, K.; Bickmore, T. Social support agents for older adults: Longitudinal affective computing in the home. *J. Multimodal User Interfaces* **2015**, *9*, 79–88. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.