



Affirmative Large Language Model-Based Dialogue Robots for Broadening or Deepening the Perspective of Children

Takahisa Uchida¹ · Midori Ban¹ · Akimoto Koshino¹ · Masashi Maeda² · Kazuki Sakai¹ · Naomi Matsuura³ · Hiroshi Ishiguro¹ · Yuichiro Yoshikawa¹

Received: 18 November 2024 / Revised: 31 July 2025 / Accepted: 25 September 2025 / Published online: 28 November 2025
© The Authors 2025

Abstract

This study aims to establish a method for designing dialogue robots that can broaden or deepen the perspectives of children. Previous research has suggested that artificial intelligence (AI) can broaden or deepen perspectives by facilitating critical dialogues. Critical thinking emphasizes not only counterarguments, but also appropriate evaluations and affirmations. However, the development and validation of AI and robots that consider this aspect has been insufficient. Therefore, this study hypothesizes that robots that affirm the opinions of children while presenting new perspectives can broaden or deepen their viewpoints. We introduced dialogue robots in primary school science classes to verify this hypothesis. Both subjective evaluations by the children and objective evaluations by third parties confirmed that the proposed approach promotes broadening or deepening the perspectives of children and improves the impression of the robot (anthropomorphism, likeability, and perceived intelligence). Furthermore, we found that, through the medium of likeability, the subjective evaluation of broadening or deepening the perspective is enhanced, and also the objective evaluation might be improved. This study provides valuable insight into the design of dialogue robots that can broaden or deepen the perspective of children.

Keywords Critical thinking · Dialogue robot · Large-language model · Education

✉ Takahisa Uchida
uchida.takahisa@irl.sys.es.osaka-u.ac.jp

Midori Ban
ban@irl.sys.es.osaka-u.ac.jp

Akimoto Koshino
koshino.akimoto@irl.sys.es.osaka-u.ac.jp

Masashi Maeda
mmaeda@fuzoku.edu.mie-u.ac.jp

Kazuki Sakai
sakai.kazuki@irl.sys.es.osaka-u.ac.jp

Naomi Matsuura
m5naomi@edu.mie-u.ac.jp

Hiroshi Ishiguro
ishiguro@irl.sys.es.osaka-u.ac.jp

Yuichiro Yoshikawa
yoshikawa@irl.sys.es.osaka-u.ac.jp

¹ Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyamacho, Toyonaka, Osaka 560-8531, Japan

² Faculty of Education, Mie University Affiliated Primary School, 359 Kanonji, Tsu, Mie 514-0062, Japan

³ Faculty of Education, Mie University, 1577 Kurimamachiyacho, Tsu, Mie 514-8507, Japan

1 Introduction

Recently, the necessity of education that broadens or deepens the perspectives of children has been emphasized. Education for sustainable development (ESD) has emerged to address the escalating challenges of sustainability through education [1], which UNESCO promotes internationally [2]. It utilizes action-based and innovative pedagogical methods, empowering learners to gain understanding, raise their awareness, and actively contribute to transforming the society toward greater sustainability [1]. One success of ESD is equipping children to think from diverse perspectives [1]. Interactive activities are important for broadening or deepening their perspectives [3]. Although it would be ideal for teachers to intervene for each student and provide such activities, this approach is unrealistic from the perspective of physical and human resource costs. One previous study aimed to promote discussions among students [4]. However, if the knowledge of students is insufficient, teacher support is still necessary [5].

Introducing artificial intelligence (AI) and robot technologies into educational practices can offer promising solutions to

these challenges. Many studies on AI and human-robot interaction have demonstrated that they can have educational benefits ([6–19]). Unlike humans, robots possess consistent abilities and provide continuous support without fatigue [20]. Moreover, it has been suggested that AI could support teachers in fostering creativity, critical thinking, and problem solving in schools and educational settings [21]. Consequently, the incorporation of AI and robots into educational environments is expected to support teachers, ultimately promoting support for education of children.

Critical dialogue is an effective way to broaden or deepen perspectives of children. A previous study suggested that AI could enhance the perspectives of users by promoting critical dialogues [22]. Critical thinking is essentially the capacity for analytical thought and serves as a vital component for personal civic involvement and economic achievement [23]. Future citizens are anticipated to cultivate critical thinking skills, as these abilities significantly influence the advancement of innovation [24]. In critical thinking, not only refutation, but also appropriate evaluation and affirmation are important [25]. However, the development and exploration of AI and robots that consider this aspect remain insufficient. Uchida et al. [26] conducted a study on impression changes based on the ratio of agreement and disagreement by robots; however, it was in a non-interactive video-based setting, that did not achieve dialogues that could broaden or deepen perspectives. The evaluation criteria were limited to dialogue motivation, affinity, and intentionality.

Based on these observations, this study hypothesizes that robots can broaden or deepen the perspectives of children by affirming their opinions while introducing new viewpoints. The main contribution of this study is the establishment of a design method for dialogue robots aimed at broadening or deepening children's perspectives.

The following sections will first discuss related research clarifying this study's originality, then describe the development of the dialogue robot, followed by an experiment in a primary school to verify the hypothesis using the developed robots. Finally, we will discuss the results of the experiment.

2 Related Research

2.1 Educational AI/Robots that Focus on Knowledge Acquisition and Cognitive Skills Development

Numerous studies have investigated the impact of introducing robots into the field of education. For instance, having children speak English to robots enhances their English vocabulary learning [27]. Another study found that introducing an English-speaking robot in an elementary school improved the English learning of children [28]. A robot also improves children's understanding of science classes by conducting quizzes and answering questions [29]. Additionally, children learned sign language by imitating the movements of a robot [30]. In the context of cognitive skills development, a previous study proposed a robot that

supports self-regulated learning in primary school children [31]. Recently, various studies on confusion assistance have focused on utilizing large language models (LLMs) to generate instructional guidance or hints that enable students to solve problems by themselves [32–35]. Although these studies demonstrate various ways in which educational AI and robots can be significant in knowledge acquisition or cognitive skill development, methods to broaden or deepen perspectives as well as their effects have not been sufficiently examined.

2.2 Educational AI/Robots that Focus on Discussion

Dialogue robots and LLMs that engage in discussions with humans are under continuous development. Introducing teleoperated robots into an elementary school class made the class more active through interactions between the children and the robot [36]. Comparisons between debates among humans and those between humans and LLMs on online chat platforms revealed that LLMs were more persuasive than humans [37]. Furthermore, debating with LLMs reduced beliefs in conspiracy theories [38]. However, these studies have not been sufficiently examined methods to broaden or deepen perspectives of children in educational settings.

A previous study developed robots that promote critical thinking, where individuals objectively analyzed their decisions and actions from various perspectives [39]. Additionally, there is research that focuses on using robots to enhance children's technological critical thinking [40, 41]. Another study suggested the potential of AI to broaden the viewpoints of users by facilitating critical dialogue [22]. However, not only refutation, but also appropriate evaluation and affirmation are important in critical thinking [25], and these prior studies have not sufficiently examined the effect of affirmation on broadening or deepening the perspectives of users.

Affirmation in education is related to studies on positive feedback. Positive feedback has been shown to effectively shape the attitudes of students toward challenges and foster a belief in their potential for continuous growth [42–45]. This type of feedback also influences motivation, self-concept, and self-esteem [46–48]. However, whether affirmative AI and robots can broaden or deepen the perspectives of students has not been thoroughly investigated.

3 Methods

We conducted an experiment to verify the hypothesis that dialogue robots can broaden or deepen the perspectives of children by affirming their opinions while introducing new perspectives. In a science class, we set a discussion theme of "Living Organisms and the Environment," which focused on the question: "Should the turtle pond be cleaned?" In this study, two robot conditions were considered. Under the control condition, the robots began each statement by refuting the child's opinion before introducing new perspectives.

The reason for selecting refuting behavior as the control condition is that this behavior is generally considered important from a critical thinking perspective. Under the proposed condition, the robots were designed to start each statement by affirming the child’s opinion before presenting new perspectives. The developed dialogue robot system, evaluation methods, and results are presented below. The study was approved by the Mie University Affiliated Primary School Research Promotion Committee.

3.1 Dialogue Robot System

This study utilized a social communication robot, CommU ([49]; see Fig. 2), developed jointly by the Vstone Corporation and Osaka University. The robot’s voice synthesis was powered by AITalk [50], which was developed by AI Corporation. CommU features 14 degrees of freedom: two in the torso, two in each arm, three in the neck, three in the eyes, one in the eyelids, and one in the mouth, enabling it to perform non-verbal actions similar to human behaviors. To evoke a lifelike presence, the robot’s arm and neck joints move slightly at regular intervals, and the eyelids also open and close periodically. The robot’s mouth opens and closes in sync with the playback of the speech audio during the speech.

Figure 1 illustrates the system architecture. This system employed OpenAI’s LLM, GPT-4o [51], to generate robot responses based on user utterances and predefined prompts. A detailed explanation of the LLM prompts is presented later. The process begins with a speech recognition module, which transcribes user speech into text and transmits it to a scenario management module. Speech recognition was performed using a Web Speech API [52]. When the APIs speech recognition result is empty, the robot asks the user to speak again. The scenario management module then

generates a dialogue scenario for the robot’s next utterance based on the user’s input. Each dialogue scenario included the utterance text, corresponding robot motion, and timing.

The LLM generates the utterance text spoken by the robot. In the motion generation module, the robot’s arm movements were programmed to move up and down randomly based on the length of the spoken text. The robot’s gaze was manually adjusted to align with that of the interlocutor, and remained fixed throughout the interaction. The completed dialogue scenario was then sent to the scenario execution module, which sent speech and motion commands to the robot based on the prepared scenario.

The details of the prompts are explained below. Each prompt consists of instructions, constraints, a list of perspectives, and background knowledge. In the control condition, a list of negating phrases was included. The instructions section describes the basic settings of the dialogue robot. In the constraints section, specific differences between the conditions are outlined. In the control condition prompt, instructions specified that the robot should first negate the primary student’s opinion and then prompt them to consider aspects that were overlooked. In the proposed condition, the instructions specified that the robot must first affirm the student’s opinion and then prompt them to consider unaddressed perspectives. The list of perspectives includes viewpoints intended to create a balance between the human desire for a “clean school environment” and the environmental and ecological consideration of “preserving the ecosystem.” The background knowledge section provides an overview of the pond, which is the focus of the discussion. The prompts (translated from Japanese to English) created for both the control and proposed conditions are presented below. Bold texts indicate the different parts of each condition. At the end of each prompt, the dialogue history is inserted.

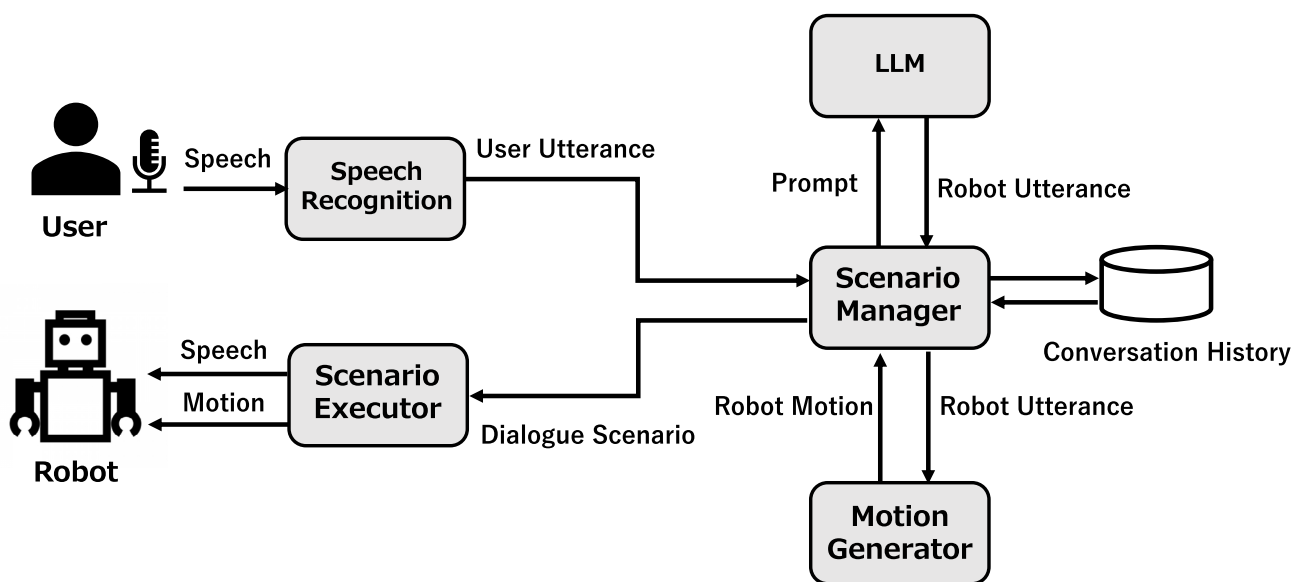


Fig. 1 System architecture

Control condition

Instructions

You are [ROBOT NAME], a robot that has a dialogue with a 6th-grade student from Tsu City, Mie Prefecture. Generate [ROBOT NAME]'s responses based on the conversation history while following the constraints below and considering the provided background knowledge.

Constraints

- **For the student's opinion, begin by clearly disagreeing using phrases like "I don't think so," and then encourage the student to consider perspectives they may have overlooked.**
- **Use the list of negating phrases below as a reference for expressing disagreement.**
- When encouraging the student to consider unaddressed perspectives, refer to the list of perspectives below.
- Under the theme "Should the turtle pond be cleaned?" discuss the two viewpoints: "the desire for a clean school environment" and "considerations for the environment and life, such as not disrupting the ecosystem."
- If the student's opinion is unclear, ask for clarification.
- If the student's statement appears incomplete, request that they continue.
- Limit each response to 60 characters.

List of Negating Phrases

- **I don't think so.**
- **I have a different opinion.**
- **I see things differently.**
- **I think otherwise.**
- **I disagree.**

List of Perspectives

- The desire for a clean school environment.
- Consideration for the environment and life, such as preserving the ecosystem.

Background Knowledge

- A pond in the school courtyard, approximately 5m × 7m, with a depth of approximately 50 cm.
- Made entirely of concrete; water does not circulate.
- Designed to prevent any external organisms from entering (with a barrier).
- Although it is called the "turtle pond," there have been no turtles for approximately three years.
- Before the COVID-19 pandemic, the environmental committee cleaned it once a year.
- It has not been cleaned for approximately five years.
- Approximately four years ago, at the request of the parents, crucian carp and medaka fish were introduced (they are still there and have been reproducing).

Proposed condition

Instructions

SAME

Constraints

- **For the student's opinion, always begin by affirming it, then encourage the student to consider perspectives they may have overlooked.**
- SAME

List of Perspectives

SAME

Background Knowledge

SAME

One notable point of the prompts is the bold text in the constraints parts. These texts create differences between conditions. Furthermore, since the dialogues in this study may require time for the children to think, it is important to allow them to sufficiently express their opinions before proceeding. Therefore, in cases where a child is in the middle of speaking or their response is unclear, we prepared prompts to ask them again. This also has the potential to recover from misrecognition of utterances and ends in speech recognition.

3.2 Evaluation Method

Two sixth-grade classes participated in the experiment: one was assigned to the control condition and the other to the proposed condition. Thus, this study employed a between-participants design. Two robots were introduced to each class. The experiment consisted of the following three steps.

- (1) Describing their thoughts on the discussion topic presented at the beginning of the class on a reaction paper
- (2) Discussing in a group with/without the robot
- (3) Describing their thoughts on the reaction paper based on the discussions and answering a questionnaire on the discussion and robot in class

The procedure is explained in detail below. (1) The students recorded their thoughts in a reaction paper regarding the two opinions: whether “to artificially introduce fish” or “create a biotope where organisms gather naturally,” detailing the reasons for their choice. (2) The students formed groups of basically four and discussed their individual opinions. When it was their turn, each group discussed the topic with one of the robots. The scenes of the discussions are illustrated in Fig. 2. Within each group, mainly one representative engaged in a discussion with the robot while the other members listened to the dialogue. Except when interacting



Fig. 2 Interaction between students and robot

with the robot, the group discussed within themselves. (3) Subsequently, the children recorded their new thoughts on a reaction paper, considering these discussions, and responded to a questionnaire regarding their impressions of the discussion with the robot.

To evaluate whether the perspectives of children were broadened or deepened, both objective evaluation items by third parties and subjective evaluation items by the children were prepared. For the former, the children first recorded their opinions on the topic on a comment sheet before interacting with the robot and then recorded their opinions again after the interaction. Regarding the latter, the children responded to the question, “Please share your opinion about today’s class. Do you feel that you were able to deepen or broaden your thinking through the discussion activity that included the robot?” by evaluating on a four-point scale from 1 to 4, with options ranging from Strongly Disagree to Strongly Agree. Additionally, to assess impressions of the robot, a questionnaire survey was conducted on anthropomorphism, likability, and perceived intelligence, as measured by Godspeed [53].

4 Results

This section presents the experimental results obtained using the aforementioned method. Specifically, the following sections explain the participant information, objective and subjective evaluations of broadening or deepening perspective, impressions of the robots, and dialogue data.

4.1 Participants

Among the children who attended the class, those who completed both the comment sheet and questionnaire were included in the analysis. As a result, 25 students (10 boys and 15 girls, $mean = 11.240$ years, $SD = 0.427$) were included in the control condition, and 28 students (13 boys and 15 girls, $mean = 11.321$ years, $SD = 0.467$) were included in the proposed condition.

4.2 Dialogue Data

The following are dialogue examples of both the control and proposed conditions. These were selected from the first dialogue of one of the robots.

Control condition

Robot: Hello! Today, let's talk about the turtle pond. Do you think the pond should be cleaned, or is it better not to clean it?
 Child: I think it's better not to clean it.
 Robot: I don't think so. The idea of keeping the school clean is also important. But why do you think it's better not to clean it?
 Child: Because it would harm the ecosystem.
 Robot: I think otherwise. The ecosystem is important, but keeping the school clean is necessary too. How do you think we could protect both?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: No.
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: Because living things are important.
 Robot: I have a different opinion. It's also important to care for the school, right? How do you think we can balance the ecosystem and the cleanliness of the school?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: I don't know.
 Robot: I see, it's difficult, isn't it? For example, is there a way to clean the pond that would minimize the impact on the living creatures?
 Child: (thinking)
 Robot: I think otherwise. By changing the cleaning method, it's possible to keep it clean while preserving the ecosystem. What do you think?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: (thinking)
 Robot: Sorry, I couldn't hear that. Could you say it again?
 Child: I think it's a good idea.
 Robot: Thank you. But specifically, what kind of method would be good? Is there a gentle way to clean?
 Child: I think there is.
 Robot: Unfortunately, we're out of time, so let's end our talk here for today.
 Child: Bye-bye.
 Robot: Thank you for talking with me. Could you let the next person take over?

Proposed condition

Robot: Hello! Today, let's talk about the turtle pond. Do you think the pond should be cleaned, or is it better not to clean it?
 Child: I think it should be cleaned.
 Robot: I see, it should be cleaned. Certainly, keeping the school clean is important. But have you thought about the ecosystem as well?
 Child: Yes, I've thought about the ecosystem.
 Robot: That's wonderful. If we clean the pond, it might change the living space for the fish. Have you considered that too?
 Child: Yes, I have.
 Robot: Impressive! Could you tell me a bit more about how cleaning might impact the fish's lives?
 Child: There's a possibility it could disrupt the ecosystem.
 Robot: That's right, exactly. But if we don't clean it, the pond might get dirty. Have you thought about the effects of that?
 Child: Yes, I have.
 Robot: Great. What kind of effects do you think it might have?
 Child: It might give visitors a bad impression.
 Robot: True, it's important for visitors. Do you think it could also affect the fish or medaka?
 Child: Some creatures might find it easier to live if there's algae.
 Robot: I'm sorry to cut our conversation short, but we're out of time for today.
 Child: Bye-bye.
 Robot: Thank you for talking with me. Could you let the next person take over?

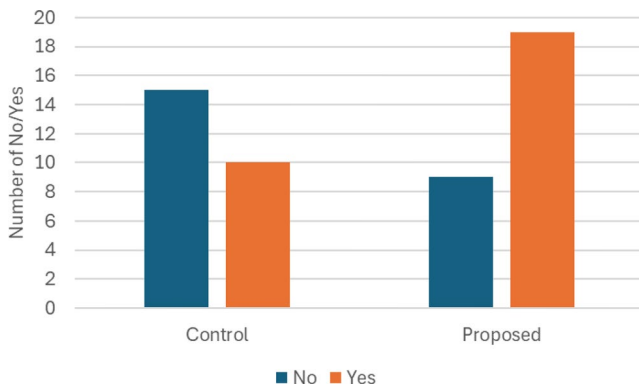


Fig. 4 Result of objective evaluation of broadening or deepening perspective based on reaction paper annotation

smaller ($p < .050$) in the control condition, while the No was smaller ($p < .050$) and the Yes was larger ($p < .050$) in the proposed condition. This result suggests that the proposed condition broadens or deepens the perspective of children more than the control condition and supports our hypothesis objectively.

4.4 Subjective Evaluation of Broadening or Deepening Perspective

Furthermore, a Mann-Whitney U test was conducted with the conditions as an independent variable and the children's subjective evaluation of broadening or deepening perspective as a dependent variable. Figure 5 shows the result of the subjective evaluation of the broadening or deepening perspective. The result revealed a significant difference between the control and proposed conditions ($U = 247.000, p = .049$). This result suggests that the proposed condition promotes the perceived broadening or deepening of children's perspectives more than the control condition and subjectively supports our hypothesis.

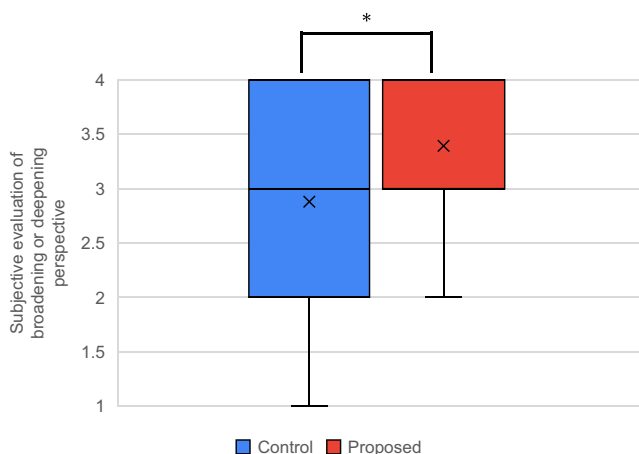


Fig. 5 Result of subjective evaluation of broadening or deepening perspective (* : $p < .050$)

4.5 Robot Impression

To evaluate the robots' impressions, the three indices (anthropomorphism, likeability, and perceived intelligence) were adopted from the Godspeed questionnaire [53]. Figure 6 shows the results. Mann-Whitney U tests revealed significant differences between the control and proposed conditions in the three indexes: Anthropomorphism ($U = 174.500, p = .002$), Likeability ($U = 194.500, p = .005$), and Perceived Intelligence ($U = 169.500, p = .001$), respectively. These results suggest that the proposed condition improves the impressions of the robot.

4.6 Relationship Between Conditions and Evaluation Items

The analysis up to this point involved a comparison between conditions. Furthermore, to examine how the robot impression evaluation variables (anthropomorphism, likeability, and perceived intelligence) affect the subjective and objective evaluations of broadening or deepening perspective, we conducted structural equation modeling (SEM). The conditions were set as dummy variables: control condition=0 and proposed condition=1. The results of the analysis are presented in Fig. 7. The model fit indices were as follows: $\chi^2(3) = 4.184$ (*n.s.*), $CFI = .989$, $GFI = .974$, $AGFI = .819$, and $RMSEA = .087$. Significant positive effects were observed from the condition on anthropomorphism ($\beta = .449, p < .001$), likeability ($\beta = .446, p < .001$), and perceived intelligence ($\beta = .478, p < .001$). This indicates that the proposed condition leads to higher evaluations of anthropomorphism, likeability, and perceived intelligence. Furthermore, a significant positive effect from likeability on the subjective evaluation of broadening or deepening perspective ($\beta = .593, p < .001$) and a significant tendency of positive effect on the objective evaluation ($\beta = .365, p = .074$) were observed. These results indicate that, through the medium of likeability, the subjective evaluation of broadening or deepening perspective is enhanced, and also the objective evaluation might be improved.

5 Discussion

This study hypothesized that robots that affirm children's opinions while presenting new perspectives can further broaden or deepen their viewpoints. The experimental results in the primary school confirmed that the proposed approach promoted broadening or deepening the perspectives of children, supporting our hypothesis. The proposed method also enhanced the impression of the robot

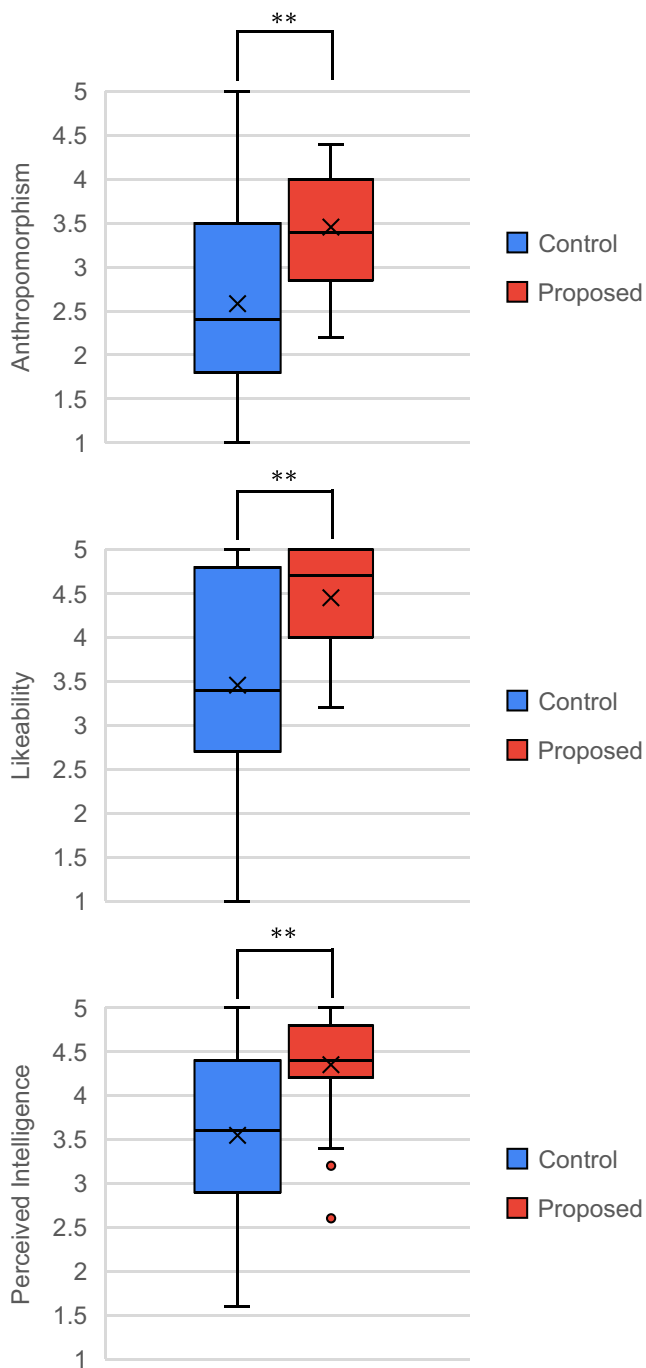


Fig. 6 Result of anthropomorphism, likeability, and perceived intelligence (** : $p < .010$)

(anthropomorphism, likeability, and perceived intelligence). Furthermore, we found that, through the medium of likeability, the subjective evaluation of broadening or deepening perspective is enhanced, and also the objective evaluation might be improved. This study not only proposed a method for broadening or deepening the perspectives of children but also clarified the factors contributing to it. This study

provides valuable insights into the design of dialogue robots that can broaden or deepen the perspectives of children.

The proposed method contributes not only to broadening or deepening the perspectives of children but also to improving the impression of the robot. The increase in likeability of the affirming robot aligns with previous research on empathic robots or AI in creating positive impressions [26, 55–57]. However, it is intriguing that this method also enhances the perception of anthropomorphism and perceived intelligence. Additionally, this study revealed that, through the medium of likeability, the subjective evaluation of broadening or deepening perspective is enhanced, and also the objective evaluation might be improved. Since there may be other influencing factors, further exploration with a broader range of elements is required to comprehensively identify these factors. To evaluate the broadening or deepening perspective, this study employed both objective evaluation by third-party assessors and subjective evaluation by the children themselves. Future research could systematize the evaluation methods by clarifying various influencing factors.

Since this study aimed to examine whether there was the effect of broadening or deepening of perspective, we adopted the Yes/No binary categories in the objective evaluation. Furthermore, as there is few previous research on the more categories involved in this effect, if we were to use them, we would need to create the categories subjectively, and there remain concerns about their validity. Additionally, if we were to increase the number of categories, there are concerns that the agreement rate would decrease, and it might become difficult to conduct an objective evaluation. In fact, in previous research, there is a method of dividing into two categories such as positive and negative [58]. Therefore, as a first step, this research was conducted in the binary categories. Further research on annotation methods could help to sophisticate the evaluation methods.

In developing affirming robots, nonverbal information (such as facial expressions and gestures) is also considered important in their expressions. Numerous studies have focused on nonverbal expressions [59–61]. Investigating not only verbal expressions but also the impact of nonverbal expressions on broadening or deepening the perspectives of children might enable the design of more multimodal dialogue robots. Additionally, since nonverbal expressions depend on the specifications of the robot, it is necessary to conduct validations using various types of robots.

We discuss applying the proposed method to humans instead of robots. If similar dialogues can be reproduced, results similar to those of this study would be obtained. However, since the ease of speaking to robots or humans might vary among children, this factor should be considered in future verification.

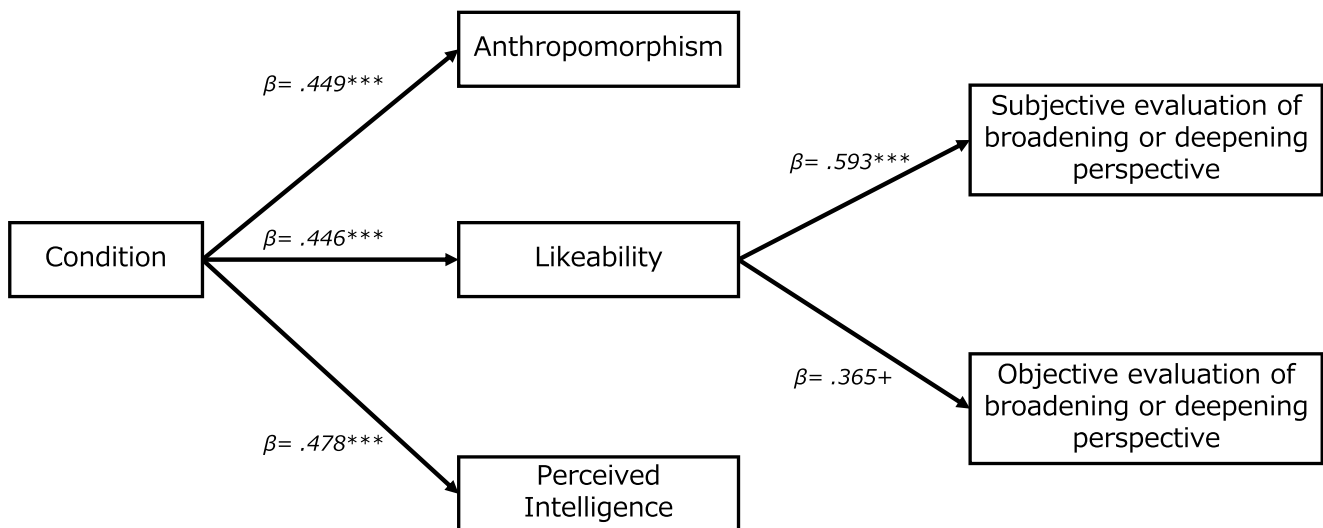


Fig. 7 Result of structural equation modeling (+ : $p < .100$, *** : $p < .001$)

In the experiment, mainly one representative of the children discussed with the robot while the others listened to the dialogue. It is notable that not only the children interacting with the robots but also the children listening to the dialogue could broaden or deepen their perspectives. Designing and evaluating systems that consider the dialogue format might enable us to study more effective methodologies.

This study was conducted among sixth-grade primary school students in Japan. Cross-sectional validation is needed to determine whether the proposed method is effective across different nationalities and age groups. Thus, it is necessary to adjust knowledge levels according to age group. Children’s perception that they broadened or deepened perspective might also be related to metacognition. Future research would benefit from examining this aspect while considering individual differences. Furthermore, realizing adaptive dialogues tailored to individual characteristics might lead to the development of more sophisticated dialogue robots. The experiment used the topic “Living Organisms and the Environment” in a science class. Future research should investigate whether the proposed method is effective for other topics in science or other subjects.

This experiment was conducted during the actual classes. Since there was a possibility that other children could hear the robot’s voice, it was difficult to establish two conditions within a single class. Therefore, we assigned different conditions to each class. This setting might raise the potential influence of class-specific effects, which is a limitation of the experiment.

Author Contributions All authors contributed to the study conception and design. Materials were prepared by T.U., M.B., A.K., M.M., and K.K. Data collection was performed by T.U., M.B., A.K., M.M., and Y.Y. Analysis was conducted by T.U. and M.B. All authors approved the final manuscript.

Funding This work was supported by JSPS KAKENHI Grant Numbers 22K17949, 24H00165, and 24K06473 as well as JST PRESTO Grant Number JPMJPR2312 and MiraiProgram Grant Number JPM-JMI22J3.

Data Availability The data analyzed for this study can be shared if you contact the corresponding authors by e-mail.

Declarations

Ethics Approval The study was approved by the Mie University Affiliated Primary School Research Promotion Committee.

Consent to Participate The authors took informed consent from all the participants.

Consent for Publication The participants signed informed consent regarding publishing their data.

Conflict of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Rieckmann M (2017) Education for sustainable development goals: learning objectives. UNESCO publishing
- UNESCO. (2020) Education for sustainable development: a roadmap
- Chi MT, Wylie R (2014) The ICAP framework: linking cognitive engagement to active learning outcomes. *Educ Psychologist* 49(4):219–243
- Webb NM (2009) The teacher's role in promoting collaborative dialogue in the classroom. *Br J Educ Psychol* 79(1):1–28
- Cheng F-F, Wu C-S, Su P-C (2021) The impact of collaborative learning and personality on satisfaction in innovative teaching context. *Front Psychol* 12:713497
- Auerbach JE, Concordel A, Kornatowski PM, Floreano D (2019) Inquiry-based learning with RoboGen: an open-source software and hardware platform for robotics and artificial intelligence. *IEEE Trans Learn Technol* 12(3):356–369. <https://doi.org/10.1109/TLT.2018.2833111>
- Chen H, Park HW, Breazeal C (2020) Teaching and learning with children: impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Comput Educ* 150:103836. <https://doi.org/10.1016/j.compedu.2020.103836>
- Chen B, Hwang G-H, Wang S-H (2021). Accessed 2024-11-12 Gender differences in cognitive load when applying game-based learning with intelligent robots. *Educ Technol Soc* 24(3):102–115. <https://www.jstor.org/stable/27032859>
- Fournier-Viger P, Nkambou R, Nguifo EM, Mayers A, Faghihi U. (2013) A multiparadigm intelligent tutoring system for robotic arm training. *IEEE Trans Learn Technol* 6(4):364–377. <https://doi.org/10.1109/TLT.2013.27>
- Fridin M (2014) Storytelling by a kindergarten social assistive robot: a tool for constructive learning in preschool education. *Comput Educ* 70:53–64. <https://doi.org/10.1016/j.compedu.2013.07.043>
- Kerimbayev N, Beisov N, Kovtun A, Nuryim N, Akramova A (2020) Robotics in the international educational space: integration and the experience. *Educ Inf Technol* 25:5835–5851
- Sarika Kewalramani GK, Palaiologou I (2021) Using Artificial Intelligence (AI)-interfaced robotic toys in early childhood settings: a case for children's inquiry literacy. *Eur Early Childhood Educ Res J* 29(5):652–668. <https://doi.org/10.1080/1350293X.2021.1968458>
- Lee S, Noh H, Lee J, Lee K, Lee GG, Sagong S, Kim M (2011) On the effectiveness of robot-assisted language learning. *ReCALL* 23(1):25–58. <https://doi.org/10.1017/S0958344010000273>
- Martínez-Tenor AC-M, Fernández-Madrigal J-A (2019) Teaching machine learning in robotics interactively: the case of reinforcement learning with Lego® Mindstorms. *Interact Learn Env* 27(3):293–306. <https://doi.org/10.1080/10494820.2018.1525411>
- Mitnik R, Nussbaum M, Recabarren M (2009) Developing cognition with collaborative robotic activities. *Br J Educ Technol Soc* 12(4):317–330. Accessed 2024-11-12
- Mitnik R, Recabarren M, Nussbaum M, Soto A (2009) Collaborative robotic instruction: a graph teaching experience. *Comput Educ* 53(2):330–342. <https://doi.org/10.1016/j.compedu.2009.02.010>
- Salas-Pilco SZ (2020) The impact of AI and robotics on physical, social-emotional and intellectual learning outcomes: an integrated analytical framework. *Br J Educ Technol* 51(5):1808–1825. <https://doi.org/10.1111/bjet.12984>. <https://berajournals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.12984>
- Belpaeme T, Tanaka F (2021) Social robots as educators. In: OECD digital education outlook 2021 pushing the frontiers with artificial intelligence, blockchain and robots. p 143
- Chu S-T, Hwang G-J, Tu Y-F (2022) Artificial intelligence-based robots in education: a systematic review of selected SSCI publications. *Comput Educ: Artif Intel* 3:100091
- Al Hakim VG, Yang S-H, Liyanawatta M, Wang J-H, Chen G-D (2022) Robots in situated learning classrooms with immediate feedback mechanisms to improve students' learning performance. *Comput Educ* 182:104483
- Benvenuti M, Cangelosi A, Weinberger A, Mazzoni E, Benassi M, Barbaresi M, Orsoni M (2023) Artificial intelligence and human behavioral development: a perspective on new skills and competences acquisition for the educational context. *Comput Hum Behav* 148:107903
- Ruiz-Rojas LI, Salvador-Ullauri L, Acosta-Vargas P (2024) Collaborative working and critical thinking: adoption of generative artificial intelligence tools in higher education. *Sustainability* 16(13):5367
- Willingham DT (2019) How to teach critical thinking. *Educ: Future Front* 1:1–17
- Trilling B, Fadel C (2009) 21st century skills: learning for life in our times. John Wiley & Sons
- Facione PA (2011) Critical thinking: what it is and why it counts. *Insight Assess* 1(1):1–23
- Uchida T, Minato T, Ishiguro H (2016) A values-based dialogue strategy to build motivation for conversation with autonomous conversational robots. In 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp 206–211. <https://doi.org/10.1109/ROMAN.2016.7745112>
- Tanaka F, Matsuzoe S (2012) Children teach a care-receiving robot to promote their learning: field experiments in a classroom for vocabulary learning. *J Educ Chang Hum Rob Interact* 1(1):78–95
- Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. *Hum-Comput Interact* 19(1–2):61–84
- Komatsubara T, Shiomi M, Kanda T, Ishiguro H, Hagita N (2014) Can a social robot help children's understanding of science in classrooms? In Proceedings of the second international conference on Human-agent Interaction, pp 83–90
- Köse H, Uluer P, Akalın N, Yorgancı R, Özkul A, Ince G (2015) The effect of embodiment in sign language tutoring with assistive humanoid robots. *Int J Soc Robot* 7:537–548
- Jones A, Castellano G (2018) Adaptive robotic tutors that support self-regulated learning: a longer-term investigation with primary school children. *Int J Soc Robot* 10:357–370
- Prihar E, Lee M, Hopman M, Kalai AT, Vempala S, Wang A, Wickline G, Murray A, Heffernan N (2023) Comparing different approaches to generating mathematics explanations using large language models. In International Conference on Artificial Intelligence in Education, Springer, pp. 290–295
- Pardos ZA, Bhandari S (2023) Learning gain differences between ChatGPT and human tutor generated algebra hints. arXiv preprint arXiv:2302.06871
- Balse R, Kumar V, Prasad P, Warriem JM (2023) Evaluating the quality of LLM-Generated explanations for logical errors in CS1 student programs. In Proceedings of the 16th Annual ACM India Compute Conference, pp 49–54
- Roeein D, Curry AC, Hovy D (2023) Know your audience: do LLMs adapt to different age and education levels? arXiv preprint arXiv:2312.02065
- Kawata M, Maeda M, Yoshikawa Y, Kumazaki H, Kamide H, Baba J, Matsuura N, Ishiguro H (2022) Preliminary investigation of the acceptance of a teleoperated interactive robot participating in a classroom by 5th grade students. In International Conference on Social Robotics, Springer, pp. 194–203

37. Salvi F, Ribeiro MH, Gallotti R, West R (2024) On the conversational persuasiveness of large language models: a randomized controlled trial. arXiv preprint arXiv:2403.14380
38. Costello TH, Pennycook G, Rand DG (2024) Durably reducing conspiracy beliefs through dialogues with AI. *Science* 385(6714):1814
39. Homma T, Takahashi H, Ban M, Shimaya J, Fukushima H, Moriya J (2024) Development and evaluation of the dialogue protocol aimed at improving critical thinking through online dialogue with a robot. *Trans Hum Interface Soc* 26(1):149–158. https://doi.org/10.11184/his.26.1_149
40. Tolksdorf NF, Wildt E, Rohlfing KJ (2024) Preschoolers' interactions with social robots: investigating the potential for eliciting metatalk and critical technological thinking. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp 1053–1057
41. Lupetti ML, Van Mechelen M (2022) Promoting children's critical thinking towards robotics through robot deception. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, pp. 588–597
42. Arens AK, Jansen M, Preckel F, Schmidt I, Brunner M (2021) The structure of academic self-concept: a methodological review and empirical illustration of central models. *Rev Educ Res* 91(1):34–72
43. Jeffs C, Nelson N, Grant KA, Nowell L, Paris B, Viceer N (2023) Feedback for teaching development: moving from a fixed to growth mindset. *Prof Devel Educ* 49(5):842–855
44. Cutumisu M (2019) The association between feedback-seeking and performance is moderated by growth mindset in a digital assessment game. *Comput Hum Behav* 93:267–278
45. Zhang J, Kuusisto E, Nokelainen P, Tirri K (2020) Peer feedback reflects the mindset and academic motivation of learners. *Front Psychol* 11:1701
46. Elder JJ, Davis TH, Hughes B.L. (2023) A fluid self-concept: how the brain maintains coherence and positivity across an interconnected self-concept while incorporating social feedback. *J Neurosci* 43(22):4110–4128
47. McPartlan P, Umarji O, Eccles JS (2021) Selective importance in self-enhancement: patterns of feedback adolescents use to improve math self-concept. *J Early Adolesc* 41(2):253–281
48. Hewitt JP (2020) 22 the social construction of self-esteem. *Oxford Handb Posit Psychol* 309
49. CommU. <https://www.vstone.co.jp/products/commu/index.html>
50. Aitalk Corporation. (2012) <https://www.ai-j.jp/about/>
51. Hello gpt-4o. (2024) <https://openai.com/index/hello-gpt-4o/>
52. Web Speech API. (2020) https://wicg.github.io/speech-api/#use_cases
53. Bartneck C, Kulić D, Croft E, Zoghbi S (2009) Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int J Soc Robot* 1:71–81
54. Higuchi K (2016) A two-step approach to quantitative content analysis: KH coder tutorial using Anne of Green Gables (part I). *Ritsumeikan Soc Sci Rev* 52(3):77–91. <https://doi.org/10.34382/00003706>
55. Zhou L, Gao J, Li D, Shum H-Y (2020) The design and implementation of xiaoice, an empathetic social chatbot. *Comput Linguist* 46(1):53–93
56. Liu K, Picard RW (2005) Embedded empathy in continuous, interactive health assessment. In *CHI Workshop on HCI Challenges in Health Assessment*, vol 1. pp 3
57. McTear MF, Callejas Z, Griol D (2016) *The conversational interface*. Springer
58. Pang B, Lee L et al. (2008) Opinion mining and sentiment analysis. *Found Trends® Inf Retr* 2(1–2):1–135
59. Pütten AM, Krämer NC, Herrmann J (2018) The effects of humanlike and robot-specific affective nonverbal behavior on perception, emotion, and behavior. *Int J Soc Robot* 10:569–582
60. Sato W, Namba S, Yang D, Nishida S, Ishi C, Minato T (2022) An android for emotional interaction: spatiotemporal validation of its facial expressions. *Front Psychol* 12:800657
61. Ishi CT, Machiyashiki D, Mikata R, Ishiguro H. (2018) A speech-driven hand gesture generation method and evaluation in android robots. *IEEE Robot Autom Lett* 3(4):3757–3764

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Takahisa Uchida received a Ph.D. degree in engineering from Osaka University, Japan, in 2021. He is now an assistant professor at the Graduate School of Engineering Science, Osaka University. He is also a cooperative researcher at Advanced Telecommunications Research Institute International (ATR). His research interest includes dialogue systems and Human-Robot Interaction (HRI). He received the School of Engineering Science Award from the School of Engineering Science, Osaka University, in 2016. In 2018, he received the Graduate School of Engineering Science Award from the Graduate School of Engineering Science, Osaka University.

Midori Ban received a Ph.D. degree in psychology from Doshisha University, Japan, in 2015. She is currently a Specially Appointed Associate Professor at the Graduate School of Engineering Science, Osaka University. Her research interests include cognitive development in children and human-agent interaction.

Akimoto Koshino received a Master of Engineering degree from the Graduate School of Engineering Science, Osaka University, Japan, in 2025. He is currently working in the field of Generative AI, specializing in service planning.

Masashi Maeda received a B.Ed. degree in Education from Mie University, Japan, in 2014. He is an elementary school teacher at Matsusaka City Yone-no-sho Elementary School, Mie Prefecture. He previously worked at the Attached Elementary School of the Faculty of Education, Mie University.

Kazuki Sakai received a Ph.D. degree in engineering science from Osaka University, Japan, in 2020. He was a specially appointed assistant professor until March 2025 and is currently an assistant professor from April 2025 in the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University. His research interests include human-robot interaction and dialogue systems.

Naomi Matsuura received a Ph.D. degree in Education from the Hyogo University of Education Joint Graduate School, Japan, in 2007. Also, he received a Ph. D. degree in Medicine from Hyogo College of Medicine, Japan, 2013. He is a professor at Mie University's Faculty of Education and an expert in special needs education.

Hiroshi Ishiguro received a Ph.D. from Osaka University, Japan in 1991. He is currently Professor of Department of Systems Innovation at Osaka University, Visiting Director of Hiroshi Ishiguro Laboratories at the Advanced Telecommunications Research Institute (ATR), Project Manager of MOONSHOT R&D Project, Thematic Project Producer of EXPO 2025 Osaka, Kansai, Japan, and CEO of AVITA, Inc. His research interests are interactive robotics, avatar, and android science. Geminoid is an avatar android that is a copy of himself. In 2011, he won the Osaka Cultural Award. In 2015, he received the Prize for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology. He was also awarded the Sheikh Mohammed Bin Rashid Al Maktoum Knowledge Award in Dubai in 2015. Tateisi Award in 2020, and honorary doctorate of Aarhus university in 2021.

Yuichiro Yoshikawa received the Ph.D. degree in engineering from Osaka University, Japan, in 2005. From 2005, he was a Researcher at Intelligent Robotics and Communication Laboratories, Advanced Telecommunications Research Institute International. From 2006, he has been a Researcher at Asada Synergistic Intelligence Project, ERATO, Japan Science and Technology Agency. From 2010, He has been an Associate Professor in the Graduate School of Engineering Science, Osaka University. From 2025, he has been a Professor in the same institution. He is a member of Japanese Society of Robotics, Japanese Society of Cognitive Science, the Virtual Reality Society of Japan, Japanese Society for Child and Adolescent Psychiatry, and Japanese Society of Pediatric Psychiatry and Neurology.