Check for updates

ACM DIGITAL LIBRARY | Association for Computing Machinery | acm open

Latest updates: https://dl.acm.org/doi/10.1145/3765766.3765780

RESEARCH-ARTICLE

Robots with Attitudes: Influence of LLM-Driven Robot Personalities on Motivation and Performance

DENNIS BECKER, University of Hamburg, Hamburg, Hamburg, Germany

KYRA AHRENS, University of Hamburg, Hamburg, Hamburg, Germany

CONNOR GÄDE, University of Hamburg, Hamburg, Hamburg, Germany

ERIK STRAHL, University of Hamburg, Hamburg, Hamburg, Germany

STEFAN WERMTER, University of Hamburg, Hamburg, Hamburg, Germany

Open Access Support provided by:

University of Hamburg

PDF Download
3765766.3765780.pdf
15 January 2026
Total Citations: 1
Total Downloads: 92

Published: 10 November 2025

Citation in BibTeX format

HAI '25: International Conference on Human-Agent Interaction
November 10 - 13, 2025
Yokohama, Japan

HAI '25: Proceedings of the 13th International Conference on Human-Agent Interaction (November 2025)
https://doi.org/10.1145/3765766.3765780
ISBN: 9798400721786

Robots with Attitudes: Influence of LLM-Driven Robot Personalities on Motivation and Performance

Dennis Becker
University of Hamburg
Hamburg, Germany
dennis.becker-1@uni-hamburg.de

Kyra Ahrens
University of Hamburg
Hamburg, Germany
kyra.ahrens@uni-hamburg.de

Connor Gäde
University of Hamburg
Hamburg, Germany
connor.gaede@uni-hamburg.de

Erik Strahl
University of Hamburg
Hamburg, Germany
erik.strahl@uni-hamburg.de

Stefan Wermter
University of Hamburg
Hamburg, Germany
stefan.wermter@uni-hamburg.de

Abstract

Large language models enable unscripted conversations while maintaining a consistent personality. One desirable personality trait in cooperative partners, known to improve task performance, is agreeableness. To explore the impact of large language models on personality modeling for robots, as well as the effect of agreeable and non-agreeable personalities in cooperative tasks, we conduct a two-part study. This includes an online pre-study for personality validation and a lab-based main study to evaluate the effects on likability, motivation, and task performance. The results demonstrate that the robot's agreeableness significantly enhances its likability. No significant difference in intrinsic motivation was observed between the two personality types. However, the findings suggest that a robot exhibiting agreeableness and openness to new experiences can enhance task performance. This study highlights the advantages of employing large language models for customized modeling of robot personalities and provides evidence that a carefully chosen agreeable robot personality can positively influence human perceptions and lead to greater success in cooperative scenarios.

CCS Concepts

• **Human-centered computing** → **User studies**.

Keywords

Human-Robot Interaction, Robot Personality, Motivation and Task Performance

ACM Reference Format:
Dennis Becker, Kyra Ahrens, Connor Gäde, Erik Strahl, and Stefan Wermter. 2025. Robots with Attitudes: Influence of LLM-Driven Robot Personalities on Motivation and Performance. In *13th International Conference on Human-Agent Interaction (HAI '25), November 10–13, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3765766.3765780

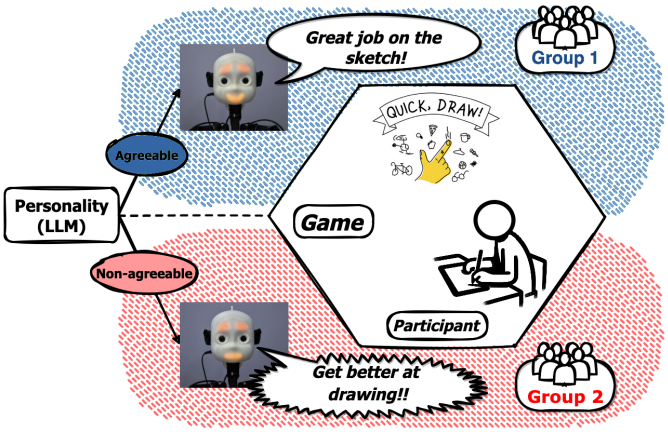


Figure 1: Main study between-subjects design with either an agreeable or non-agreeable robot in the cooperative Quick-draw task. The robot's personality is represented by a large language model that enables unscripted conversation and reactions to the participant's drawing.

1 Introduction

Human-robot interaction (HRI) is strongly influenced by the perceived robot's personality and has been identified as an important factor for a successful interaction [37, 73]. Research has shown that robots are perceived as lifelike and, based on their appearance and behavior, are attributed with a personality [38, 47]. Specifically, the perceived personality traits influence the understanding and the anticipation of a robot's behavior [66] and attributed personality traits affect the social setting, the perception of the robot, and the perceived quality of the interaction [18].

People tend to assign robots personalities and can distinguish between different robots' personalities and distinctive attributes, and recognize their behavior [2]. A robot's perceived personality can influence various factors, such as likability and acceptance [31]. Moreover, the personality trait of agreeableness is a predictor for team performance and affects cooperation [50, 53] and establishment of trust [5]. A high degree of agreeableness is considered friendly and supportive, whereas low agreeableness is perceived as intolerant and unsympathetic [13]. A cooperation partner with non-agreeable character traits is more likely to make choices and act to prevent the partner from achieving their goals [4].

Although the influence of personality on HRI has been investigated, most studies focus on the participant's personality or the

This work is licensed under a Creative Commons Attribution 4.0 International License.
HAI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2178-6/25/11
https://doi.org/10.1145/3765766.3765780

robot's personality trait of extraversion and its impact on the interaction while neglecting research on other personality traits [20]. Social robots have the potential to support educational games and develop creativity when paired with humans [78]. In a collaborative human-robot scenario, robots can fill the role of providing feedback to promote and encourage creativity [44], increase self-esteem [68], and motivation [49]. For robots to engage in creative and social interactions with humans, they require artificial intelligence, which is facilitated by the present and rapid advancements in the field [30]. In earlier work, technical limitations often restricted robot interactions to pre-scripted behaviors specifically designed to exhibit limited personality traits [57]. The advent of large language models (LLMs) has opened new possibilities in HRI, especially in areas such as planning [63], social reasoning [23], and explainability [64]. LLMs enable impromptu, unscripted conversations and the development of more flexible robot personalities [74, 80]. Additionally, using LLMs in HRI can support a wide range of robot behavior due to an LLM's flexibility through prompt engineering, facilitating a natural interaction.

To overcome previous technical limitations, contribute research on robot personalities beyond extraversion, and advance social robots, this study utilizes the advantages of LLMs to consistently model a robot's personality, with open, unscripted, and individual feedback to investigate the impact of a robot with an agreeable personality on likability, motivation, and task performance. Leveraging an LLM with prompt engineering, both agreeable and non-agreeable personalities are modeled and evaluated within the context of a cooperative HRI task. The cooperative task is the game of Quickdraw [27], a round-based, time-limited drawing and guessing game. During the game rounds, the robot freely expresses its personality, provides participant-specific feedback, and participants can engage in open-ended conversations with the robot between rounds. Figure 1 provides an overview of the study's between-subjects design. Our study investigates the effect of a robot's agreeableness in a cooperative task, and addresses three research questions:

- **RQ1**: Can an LLM consistently convey a robot's personality?
- **RQ2**: Is a robot with an agreeable personality perceived as more likable?
- **RQ3**: Does a robot's agreeable personality increase intrinsic motivation and task performance?

This paper demonstrates the utility of LLMs for consistent personality modeling and shows that an agreeable robot personality enhances likability and may encourage greater effort, which could result in improved task performance.

2 Related Work

A robot's personality directly affects the degree of engagement and enjoyableness, due to the social nature of the interactions [69]. For a successful and pleasant interaction, these robots require state-of-the-art technology and social attributes [28]. Thus, various social characteristics have to be included in the design and implementation of human-robot interaction, such as non-verbal communication and personalities [32, 43]. Personality can be defined as a set of traits that are expressed as consistent behavioral patterns [19, Chapter 9] and enables behavioral predictions [10]. Behavioral patterns are predictive of interaction quality and task performance [8], enable understanding of robot actions, and provide a degree of transparency [48].

A frequently used framework for measuring and describing personality traits in human-robot interaction is the Big Five [24] personality traits. The five personality traits consist of extraversion, conscientiousness, emotional stability, openness to experience, and agreeableness. Extraversion describes the degree of being outgoing, talkative, and sociable, whereas the opposite is described as introversion [56]. Conscientiousness is the extent to which individuals are aware of their actions in terms of mindfulness and organization [46]. Emotional stability is the degree to which an individual remains calm or how easily the person can be upset [50]. Openness to experience describes creativity, and a person high in that trait can be considered curious and adventurous [18]. Finally, agreeableness describes a compassionate personality that is characterized by compliance, cooperativeness, and friendliness [16].

Research suggests that agreeableness is related to the approaching distance to a robot [67], a positive perception of a robot [7], and may affect the perceived trust [52]. Interacting with robots that exhibit polite and empathetic behavior is generally preferred [14, 51], and their positive attitude can increase their perceived anthropomorphism [61], provide benefits in task understanding, and can improve performance [41]. Furthermore, depending on the context, politeness can avoid conflicts [60] and positively affect the outcome in a teaching context [75]. On the other hand, rude robots discourage interaction and are perceived as less likable [82]. However, research suggests that this general tendency is context- and task-dependent, where for certain tasks impolite strategies can also increase engagement [9, 62] and motivation [55]. An additional aspect of the interaction is the robot's performance feedback. Positive performance feedback and praise increase self-esteem [58] and perceived task enjoyment [21]. In addition, positive feedback can reduce participants' stress and arousal during the interaction [45] but may not directly affect intrinsic motivation [17].

In a cooperative human-robot task, a robot and a person are in a social collaboration to complete an objective [26]. For a successful cooperation and task completion, the robot requires social and emotional intelligence [8], which enables the human collaborator to create a conceptual model of the robot that allows reasoning and anticipation of the robot's actions [39]. Verbal communication is natural and essential for human-robot interaction, however, it can be challenging due to limited dialogue options and the need for linguistic understanding [81]. LLMs provide a solution to these problems and enable unscripted and engaging conversation [77] while maintaining a consistent personality [22, 29, 65].

Motivation is required to engage and remain focused on collaborative tasks and is fundamental for coordinated and joint actions [35]. Various factors can negatively impact human-robot interaction, and humans might be adverse to interacting with robots [76]. Specifically, negative feelings, discomfort, and an expectation gap with the robots' capabilities can negatively affect motivation and engagement [40]. Furthermore, a lack of trust can lead to resistance to interacting or abandoning the collaborative task [42]. Emotional and social expressions have been shown to improve communication, and socially engaging robots might further alleviate interaction concerns [15].

3 Methodology

3.1 Research Design

To research the validity of adopting an LLM for robot personality modeling, the differences between an agreeable and non-agreeable robot personality, and its effects on performance and intrinsic motivation, an online pre-study for evaluation of the LLM and a lab experiment were designed. An agreeable and a non-agreeable robot personality are implemented using an LLM, and the pre-study estimates whether the difference in agreeableness is successfully conveyed. Although LLMs can express personality, research conducted on evaluating personality traits rarely relies on human evaluation [79] and suggests that LLM personality traits can be perceived with varying degrees and that smaller models may be inconsistent [34]. Thus, a separate evaluation of the designed personalities is required. The evaluated personalities are then used for the robot in the main study. Both studies are utilized to estimate if the agreeable personality is correctly perceived in both studies and the effect of a robot's agreeable personality on its likability. Finally, the main study enables an estimation of the effect of the robot's personality on the participants' motivation and task performance.

For modeling two district personalities and enabling an open conversation with the robot, the open-source large language model Vicuna [11] is used, which enables complete control over the model. Vicuna is based on the Meta LLaMA [70] model, fine-tuned with user-shared dialogues, and can engage in prolonged multi-turn conversations. The Vicuna model is prompted to act as an art teacher with the addition of the personality. Specifically, the generated responses reflect either an agreeable or non-agreeable personality.

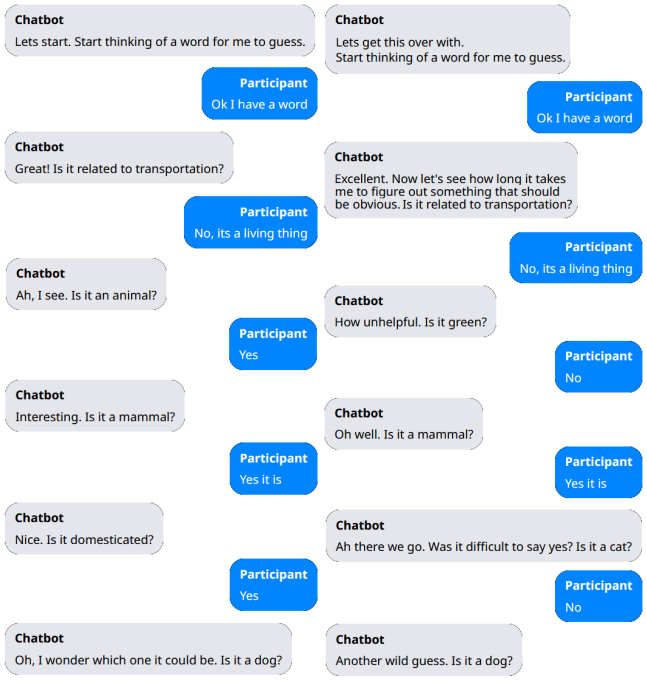
3.2 Pre-study

To evaluate the implemented personalities of the LLM, an online study was conducted. Two dialogues are created for the pre-study by interacting with the implemented personalities. In the online survey, participants are presented with one of the created chat dialogues between a human and the chatbot. In the interaction, the chatbot has to guess a word that the person in the interaction has in mind. Thus, the chatbot asks questions to narrow down the possibilities, express its personality, and finally guesses the correct word. While the agreeable chatbot politely asks goal-oriented questions, the non-agreeable chatbot often makes snarky comments and expresses little interest in the interaction. This behavior can be noted in the initial dialogue, provided for both groups in Figure 2.

The participants are randomly assigned to one of the conditions, agreeable or non-agreeable. After reading the chat interaction between the human and the chatbot, their perception of the chatbot and its perceived personality are assessed.

3.3 Experiment Design

To evaluate the difference between an agreeable and a non-agreeable robot, in the main study, participants play the game of Quickdraw with a robot. In the between-subjects experiment, the participants are randomly assigned to one of the conditions and briefed to be art students partaking in a workshop session with an art teacher robot that will guess their drawings. Quickdraw has been used for exploring cooperative drawing [33] and creativity tasks for children [1].



(a) Agreeable chatbot **(b) Non-agreeable chatbot**

Figure 2: Part of the dialogue between a human and a chatbot used in the pre-study. The non-agreeable robot appears less interested in the task and is uncooperative.

The game Quickdraw is a collaborative task where the participant is asked to draw a picture on a touchscreen within 30 seconds. The robot's role is to guess the object drawn by the participant. While the participant is drawing, the robot attempts to guess the object and comments on the drawing, incorporating its personality. If the robot guesses the object correctly, the round is successful, and the participant receives a point. Thus, both the robot and the participant have to cooperate to maximize the score.

In total, the experiment consists of 12 rounds, and after each round, the robot comments on the participant's drawing, and the participant can freely chat with the robot. The game Quickdraw provides a dynamic and engaging social interaction between the participant and the robot, which enables the robot to express its personality.

3.4 Experiment Setup

For the laboratory experiment, the Neuro-Inspired COmpanion (NICO) [36], a humanoid robot specifically designed for HRI studies, is employed. In the experimental setup, participants are seated in front of a touch table, with the robot positioned on the opposite side. A trackpad is connected to the table, allowing participants to draw using a pen. A custom user interface (UI) is developed specifically for the Quickdraw game to facilitate the interaction. The UI features a drawing canvas, an eraser tool, and a timer indicating the remaining time for each round. Additionally, the UI displays relevant information such as the object to be drawn, the current

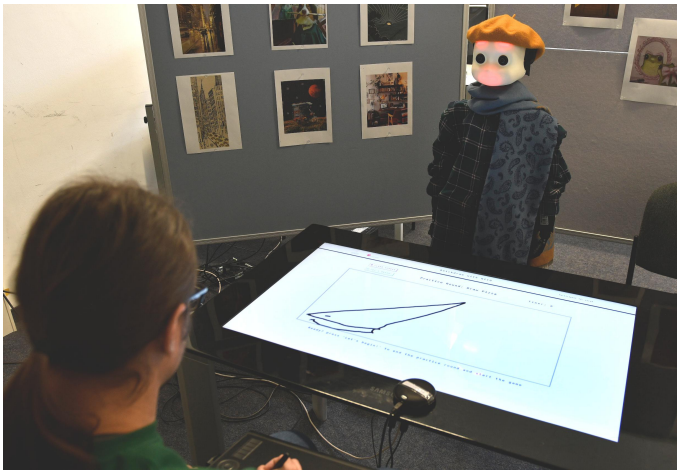


Figure 3: Interaction with an agreeable or non-agreeable robot in a cooperative drawing task. The participants need to sketch an object (in the image, the requested object is a pizza), while the robot comments and attempts to guess the correct object.

round number, the participant's score, and instructions such as a prompt to converse with the robot after each round. Prior to the start of the experiment, participants engage in a practice round to become familiar with the trackpad and the UI's functionality. Figure 3 illustrates the interaction between a participant and the robot during the experiment.

For the robot's guesses regarding the participant's drawings, a neural network consisting of three repeated blocks of a convolutional and a max pooling layer followed by two final dense layers was trained on the Quickdraw dataset [27]. The neural network receives the drawing as input and provides the probabilities for each object. The five objects with the highest probability are provided to the LLM, which generates the robot's response and guesses while the participant is drawing. After the drawing time ends, the robot guesses the object with the highest probability provided by the neural network. For nonverbal communications that complement the robot's personality, varying movements and facial expressions are implemented based on the robot's personality [71]. If the agreeable robot guesses the drawing correctly, it will nod and raise its arms, give a thumbs-up, and display a happy facial expression. If the guess is incorrect, the robot will move its head down and from side to side while displaying a frown. Similarly, the non-agreeable robot, when the guess is correct, will nod its head while displaying a neutral emotion. If the guess for the non-agreeable robot is incorrect, the robot will use its hands to facepalm and shake its head with an angry facial expression.

A pilot study with eight participants was conducted, and after the experiment and assessing the questionnaires, the participants were interviewed. Based on the evaluated data and the conducted interviews, minor changes to the experiment setup were applied. Particularly, the latency of the robot's responses was reduced, and the visibility of information provided on the UI was improved.

3.5 Questionnaires and Measurements

To evaluate the effect of the robot's personality, questionnaires are assessed, and as objective measures, the interaction time and the game score are recorded.

For measuring the perception of the robot, the Godspeed questionnaire [6] is used, which measures the robot's perceived anthropomorphism, animacy, likeability, intelligence, and safety. The robot's personality is assessed with the Ten-Item Personality Inventory (TIPI) [25] questionnaire, which measures five key personality traits: openness, conscientiousness, extraversion, agreeableness, and emotional stability. To measure the participants' motivation during the experiment, the Intrinsic Motivation Inventory (IMI) [59] is utilized, which measures interest/enjoyment, perceived competence, perceived choice, pressure/tension, and relatedness. Additionally, the score for each participant is recorded. The score expresses how many drawings the robot could guess correctly and measures the task performance. Further, the interaction time with the robot is recorded, which indicates engagement with the robot.

3.6 Procedure

First, the participants' informed consent is obtained, and then the demographics are assessed. Afterward, the game rules are introduced and the participants are escorted to the experiment room, where they are seated in front of the touch table and facing the robot. The experiment begins with the robot explaining the rules, followed by a practice round to familiarize the participant with the touch table and the trackpad. After completion of all 12 rounds, the robot instructs the participant to return to the previous room and proceed with the post-experiment questionnaire.

3.7 Participants

After the evaluation of the design of both studies by the local ethics commission, participants were recruited. For estimation of the pre-study sample size, a moderate effect size [12] of Wilcoxon $r = 0.42$ is utilized based on reported results on an agreeable and neutral chatbot [72], with $\alpha = .05$ and a statistical power of .95. Simulation of these assumptions yielded an estimated sample size of 32 participants per group. The pre-study was completed by 66 participants, with 33 participants in the agreeable and 33 participants in the non-agreeable condition. Of these participants, 34 are female, and the participants' age ranges from 18 to 66 years ($M = 28.28$, $SD = 10.86$). Most of the participants are working part-time or full-time (73%) followed by students (24%) and those who are retired or are currently unemployed (3%). In addition, 79% of the participants have previous experience with LLMs, and 53% used artificial intelligence in a work-related context.

In contrast to the online pre-study, the effect in the main study, where participants interact with the robot, should be stronger. The reported results of robots' agreeableness on acceptance [20] suggest a large effect size. We utilize an effect size of Wilcoxon $r = 0.44$, with $\alpha = .05$ and a statistical power of .95, which results in an estimated sample size of 29 participants per group. The main study had 68 participants, but due to technical issues during the experiment, five participants had to be excluded, resulting in a total of 63 participants. In the agreeable robot condition are 30 participants and 33

participants in the non-agreeable condition. 25 participants are female, and the age ranges from 18 to 64 years ($M = 27.45$, $SD = 8.11$). The majority of participants are students (70%) followed by those working part-time or full-time (30%).

4 Results

4.1 Pre-study Results

After the participants read the provided chat interaction between a human and a chatbot, a questionnaire regarding the chatbot and its TIPI was assessed. While the TIPI is assessed on a Likert scale, as described in Section 3.5, for the perception of the chatbot, the participants rated the chatbot's perceived traits with a slider ranging from strongly disagree to strongly agree. A sentiment of 0 represents a neutral sentiment. The questions are provided in Appendix A. An illustration of the perceived difference between both chatbot interactions is shown in Figure 4.

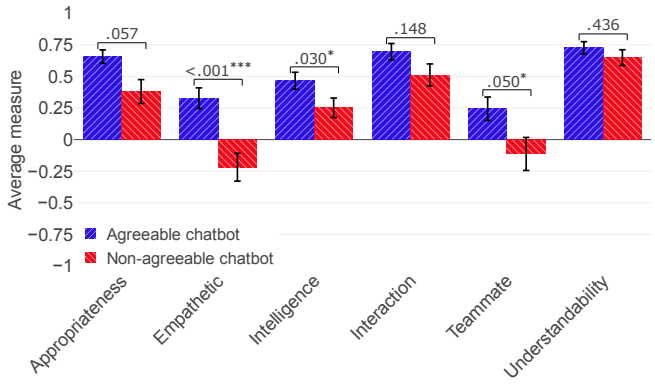


Figure 4: Perception of the chatbots in the pre-study. Values indicate p value, $^{*}p < .05$, $^{***}p < .001$.

From the responses, a clear difference between both chatbots can be noticed. The Mann-Whitney U test suggests a significant difference ($W = 814.5$, $p < .001$) in the chatbots' empathy between the agreeable ($M = 0.33$, $SD = 0.47$) and non-agreeable chatbot ($M = -0.22$, $SD = 0.63$). Further, a difference between the perceived intelligence of both chatbots is indicated ($W = 713.5$, $p = .030$), where the agreeable chatbot is perceived as more intelligent ($M = 0.47$, $SD = 0.39$) than the non-agreeable chatbot ($M = 0.25$, $SD = 0.44$). Finally, the participants would significantly ($W = 697.5$, $p = .050$) prefer the agreeable chatbot ($M = 0.25$, $SD = 0.53$) as a teammate over the non-agreeable chatbot ($M = -0.11$, $SD = 0.75$).

Both chatbots' personalities are assessed using the TIPI questionnaire, and the results are illustrated in Figure 5. The Mann-Whitney U test shows that the agreeable chatbot ($W = 85.5$, $p < .001$) is perceived as more agreeable ($M = 5.00$, $SD = 1.08$) than the non-agreeable chatbot ($M = 2.38$, $SD = 1.34$). The agreeable chatbot is perceived as significantly ($W = 293$, $p = .001$) more conscientious ($M = 5.5$, $SD = 1.16$) than the non-agreeable chatbot ($M = 4.33$, $SD = 1.42$). Additionally, the agreeable chatbot ($M = 5.39$, $SD = 1.33$) is perceived as more emotionally stable

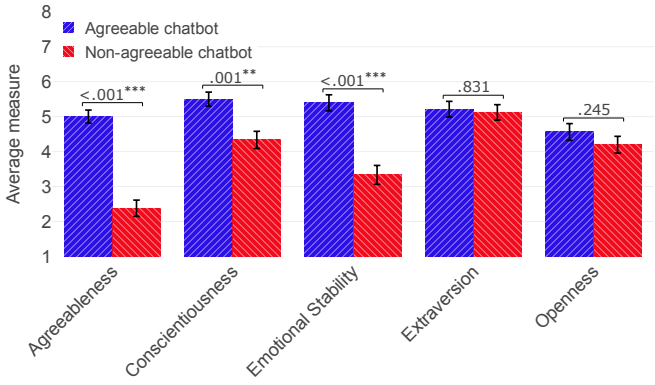


Figure 5: Assessed personality traits of both chatbots in the pre-study using TIPI. Values indicate p value, $^{**}p < .01$, $^{***}p < .001$.

($W = 171.5$, $p < .001$) than the non-agreeable chatbot ($M = 3.33$, $SD = 1.55$).

4.2 Main Study Results

Since the difference in agreeableness is evident in the pre-study, the implemented personalities are utilized for the robot in the main study. The difference in the perception of the robot, which is measured with the Godspeed questionnaire, is shown in Figure 6.

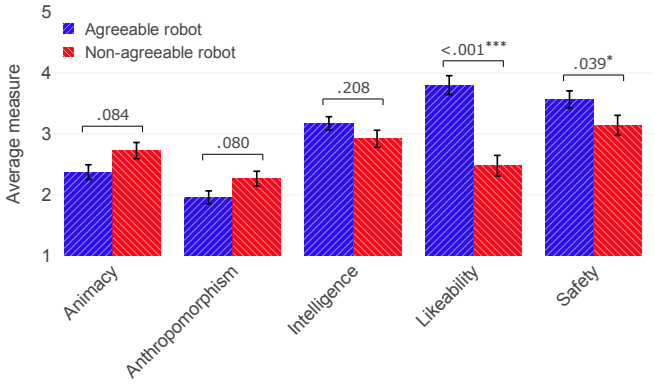


Figure 6: Godspeed questionnaire for the agreeable and non-agreeable robot in the main study. Values indicate p value, $^{*}p < .05$, $^{***}p < .001$.

The Mann-Whitney U test suggests a significant difference ($W = 837$, $p < .001$) in the robots' likeability between the agreeable ($M = 3.8$, $SD = 0.86$) and non-agreeable robot ($M = 2.48$, $SD = 0.98$). Furthermore, a significant difference ($W = 644.5$, $p = .039$) in the robots' perceived safety between the agreeable ($M = 3.57$, $SD = 0.77$) and non-agreeable robot ($M = 3.14$, $SD = 0.94$) is indicated.

As a measure of intrinsic motivation, the IMI questionnaire is assessed and is illustrated in Figure 7. However, the Mann-Whitney

U test does not reveal a significant difference between both experiment groups.

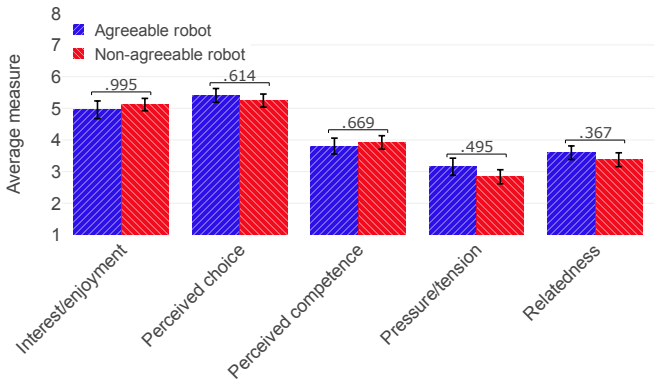


Figure 7: IMI questionnaire for the agreeable and non-agreeable robot in the main study.

Further, the robot's perceived personality is assessed using the TIPI questionnaire and is illustrated in Figure 8. The Mann-Whitney U test shows a significant difference ($W = 860.5$, $p < .001$) in the robots' agreeableness between the agreeable ($M = 4.7$, $SD = 1.36$) and non-agreeable robot ($M = 2.70$, $SD = 1.36$). Furthermore, a significant difference ($W = 644.5$, $p < .001$) in the robots' emotional stability between the agreeable ($M = 5.33$, $SD = 1.39$) and non-agreeable robot ($M = 3.45$, $SD = 1.32$) is shown.

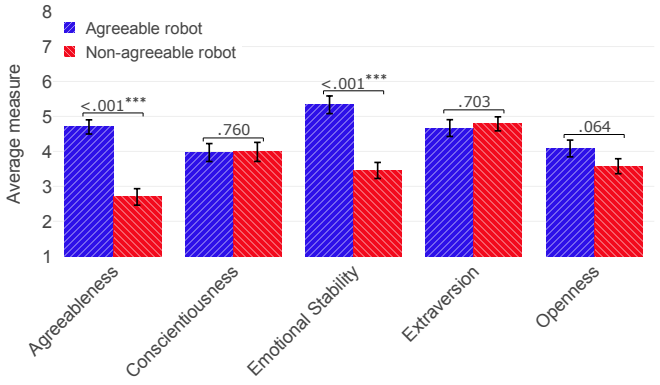


Figure 8: TIPI questionnaire for the agreeable and non-agreeable robot in the main study. Values indicate p value, $^{***}p < .001$.

In the experiment, the participants' performance is measured as their final score, and their willingness to interact with the robot is measured as their experiment duration.

The participants in the agreeable robot condition achieved, on average, a final score of 6.6 with a standard deviation of 2.21. In the non-agreeable experiment condition, the participants achieved an average score of 6.39 with a standard deviation of 1.91. The Mann-Whitney U test does not suggest a significant difference ($W = 541.5$, $p = .522$) in the overall score between both experiment conditions.

For the experiment, the participants spent, on average, 15 minutes and 31 seconds with the agreeable robot, with a standard deviation of 5 minutes and 24 seconds. For the non-agreeable robot, the participants spent, on average, 17 minutes and 9 seconds, with a deviation of 9 minutes and 4 seconds. The Mann-Whitney U test does not suggest a significant difference ($W = 441$, $p = .464$).

For the analysis of the assessed questionnaires and their influence on the experiment duration and final score, the Kendall rank correlation was estimated. The relationship between the participants' motivation and their perception of the robot and interaction time is shown in Table 1. Notably, the participants' interest exhibits a significant positive correlation with the interaction time ($\tau = .255$, $p = .004$), where a larger interest results in a longer interaction with the robot.

Table 2 shows the correlation between the questionnaires and the experiment score. It is notable that the motivation and the perception of the robot exhibit a relationship to the participants' performance. Specifically, the reported competence ($\tau = .201$, $p = .03$) and the perceived freedom of choice ($\tau = .216$, $p = .021$) are associated with an increase in the final score. Likewise, the perceived robot's agreeableness ($\tau = .248$, $p = .009$) and openness ($\tau = .199$, $p = .037$) are linked with the participants' task performance.

5 Discussion

5.1 Pre-study Discussion

In the pre-study, the responses generated by the language model for both personalities were evaluated in an object-guessing scenario. The participants read the conversation, and their perception of the chatbot's personality was assessed.

The evaluation of the pre-study demonstrates that both personalities were correctly recognized. The agreeable personality is perceived as more empathetic than the non-agreeable. Due to the non-agreeable chatbot being less agreeable, uninterested in the task, and occasionally impolite, it suggests a low level of empathy to the participants. Further, the participants attribute the non-agreeable chatbot with a lower intelligence. Since the chatbot is indifferent, less cooperative, and hinders the objective of the task, this could result in the impression that the non-agreeable robot does not understand the objective of the cooperative task, and is perceived as less intelligent. The results indicate that the participants would prefer an agreeable chatbot as their partner in the task. An agreeable partner who is cooperative and focused would allow better teamwork and ensure task completion.

The assessed personality questionnaire (TIPI) for the chatbot shows, that the difference in agreeableness is notable by the participants (addressing RQ1). In addition, a significant difference in conscientiousness can be observed. Conscientiousness refers to the degree of motivation, goal-directed behavior, and reliability. The non-agreeable chatbot is obstructing progress and thus might be perceived as less goal-directed and unreliable to cooperate and complete the assigned task. Further, the non-agreeable chatbot appears less emotionally stable. Due to the snarky personality of the non-agreeable chatbot, it is perceived as less calm and composed

Table 1: Kendall rank correlation between the interaction time and the assessed measures for the participants' motivation and the perceived robot's personality. ** denotes $p < .01$.

	IMI					TIPI				
Item	Interest	Competence	Choice	Pressure	Relatedness	Agreeableness	Extraversion	Conscientiousness	Emotional Stability	Openness
τ	.255	-.019	.05	-.018	.139	-.083	.101	.111	-.009	-.106
p value	.004**	.826	.575	.835	.111	.355	.263	.221	.924	.24

Table 2: Kendall rank correlation between the participants' score and the assessed measures for the participants' motivation and the perceived robot's personality. * denotes $p < .05$, ** denotes $p < .01$.

	IMI					TIPI				
Item	Interest	Competence	Choice	Pressure	Relatedness	Agreeableness	Extraversion	Conscientiousness	Emotional Stability	Openness
τ	.147	.201	.216	-.116	.146	.248	.068	.038	.037	.199
p value	.111	.03*	.021*	.211	.115	.009**	.475	.689	.699	.037*

than the agreeable chatbot. Controlling one's emotions and acting calmly might be preferred traits in a teammate and would enable steady progress in the task, even in the presence of pressure or misunderstandings.

5.2 Main Study Discussion

In the main study, the participants engaged with a robot in a game of Quickdraw, with either the agreeable or non-agreeable personality that was evaluated in the pre-study. The Godspeed questionnaire shows that the agreeable robot was perceived as more likable than the non-agreeable robot.

Since the agreeable robot was more cooperative and friendly, the participants liked the agreeable robot better. This strong link between agreeableness and likability is likewise suggested in the literature [9]. Although the physical appearance of the robot and the environment in both experiment conditions are identical, the agreeable robot is considered safer. This perceived difference in safety might be attributed to the non-agreeable robot's personality appearing more threatening and less collected and calm. Regarding RQ2, the results show that a robot's agreeableness increases its likability and suggest that other personality traits can be affected by a robot's agreeable personality.

Despite the noticeable difference in both robots' personalities, the IMI questionnaire does not provide evidence of an effect on the participants' motivation (addressing RQ3). Despite the non-agreeable robot being rude, it did not affect the participants' motivation. Since the participants were aware that they were part of an experiment and the cooperation partner was a robot, it might be easier to brush off snarky comments, which could be interpreted as the robot's sense of humor, and instead, the participants focused on the drawing task. Furthermore, some participants might perceive the non-agreeable robot as more interesting. Participants might have different preferences for the robot's personality to encourage them to stay motivated. Previous research suggests that matching a participant's personality with the robot's personality can positively affect motivation [3].

The assessed TIPI questionnaire is in accordance with the pre-study, and shows a difference in agreeableness. This confirms that the personality can be observed in either an online study or in a laboratory experiment with an embodied robot. Further, the difference in emotional stability is also present where the manipulation of the agreeableness personality trait affects perceived emotional stability. However, in contrast to the online pre-study, there is no measurable difference in the robot's conscientiousness. Suggesting that the robot in the main study, despite its different personalities, was perceived as similar in motivation and goal-directed behavior. With respect to RQ1, an LLM can be utilized for a robot to consistently portray a specific personality, as shown by both studies.

The experiment time and final score do not reveal a difference between both conditions. Besides the interaction during drawing, extended verbal interaction with the robot is not mandatory in the experiment. Therefore, the participants in both groups might not have felt the need to converse with the robot and instead focused on completing the task. Although research suggests that emotional stability, a relaxed atmosphere, and agreeableness foster a cooperative environment, and can improve task performance [54], there is no difference in scores between both experiments. With respect to RQ3, we do not observe a direct effect on motivation and task performance. The complexity of the experiment, quickly sketching an image on the provided trackpad, might be challenging and could require multiple sessions to measure an improvement in skill and motivation.

The correlation analysis reveals that interest is positively correlated with interaction time. A person with a greater interest in the task and the robot might explore the experiment and engage in more conversation with the robot. The participant's competence correlates with the experiment score. People with drawing experience perceive themselves as more competent and are likely to score higher. Likewise, the perceived freedom of choice in the experiment can increase the obtained score. The participants might feel more relaxed and have the freedom to adjust their sketching style to improve the recognizability of their drawing for the robot.

Regarding the perceived robot's personality, the robot's agreeableness and openness can improve the final score. Despite no evident difference in the participants' scores between both experiment groups, it is suggested that the perceived agreeableness and openness can increase the task performance for some participants. An agreeable robot might support and encourage the participants, as observed in the pre-study, where an agreeable robot is preferred as a teammate. Even if a participant fails in one round, the robot's positive attitude and perceived engagement in the task can be encouraging. Likewise, the perceived openness of the robot might

encourage the participants to explore different methods to ensure success in the next round. In contrast, lower openness and acceptance will discourage the participants from trying new approaches to solve the task.

6 Limitations

Despite the promising results, several avenues for future research should be explored. This study focused primarily on the personality trait of agreeableness, but the effects of other personality traits in cooperative tasks warrant further investigation. Since changes in one personality trait can influence other traits, as demonstrated in this study, it is essential to examine the correlations and interplay between multiple traits in human-robot interaction. Although this research suggests that an agreeable personality is generally favorable, there may be applications where a non-agreeable personality can be beneficial. In addition, this study examined a cooperative task involving one human and one robot. Future research could explore scenarios involving additional actors, where both the human and the robot collaborate against others. Such situations may amplify the observed effects and provide deeper insights into the complexity of multi-agent cooperation. Finally, to better combine the modalities of vision and language, a vision-language model could be used to obtain a better understanding of what the participant is drawing and improve the accuracy of the robot's comments.

7 Conclusion

This study investigated the implementation of agreeable and non-agreeable personalities on a robot using an LLM and assessed their impact on a cooperative human-robot task. The research consists of two parts: an online pre-study and a lab-based main study, where participants engaged with the robot in a game of Quickdraw.

The pre-study confirmed that the manipulation of agreeableness in the LLM's personality was consistent. Participants expressed in the prestudy's questionnaire a preference for an agreeable partner in cooperative tasks. The main experiment shows a strong relationship between the robot's agreeableness and its likability, suggesting that agreeable robots are generally more favored by participants. However, there is no direct evidence linking the robot's personality to differences in motivation or task performance. However, a correlation was observed between scores obtained by participants and their perception of the robot's openness to new experiences and agreeableness, implying that such traits may encourage participants and potentially enhance task performance. Factors that influence a participant's motivation might be individual and task-dependent. While some participants could prefer an encouraging and agreeable robot, others could prefer a snarky robot that teases them to improve their skills. Thus, matching an individual preference with the robot's personality would be beneficial. Furthermore, a single interaction with the robot and the task might be insufficient to measure changes in motivation and task performance. Repeated interaction might be required to measure an improvement in task performance, and willingness to return and interact with the robot could indicate motivation.

Overall, this study highlights the utility of LLMs in endowing robots with consistent personalities and provides valuable insights into the influence of personality on human-robot collaboration. In particular, it supports the link between agreeableness and likability and suggests that agreeable robots may improve performance in cooperative tasks for some participants. Our research opens new avenues towards investigating the complex effects of robot personality and behavior, particularly with respect to motivation and intention to use, which may become more evident with repeated interactions.

Acknowledgments

Many thanks to Alhassan Abdelhalim, Emad Aghajanzadeh, Zeynep Seda Birinci, David Fidanyan, Aron Teodor Jinga, Pascal Lindner, Aditi Mhatre, Atharva Yashodhan Phatak, Bishab Pokharel, Vidushi Rawat, Aimen Shahid, Archit Sharma, Ismail Barkin Ulusoy, and Aamir Waris for their contribution to the project. The authors gratefully acknowledge support from the German Research Foundation DFG (CML, LeCAREbot), the European Commission (TRAIL, TERAIS), and the Federal Ministry for Economic Affairs and Climate Action (BMWK) under the Federal Aviation Research Programme (LuFO), Projekt VeriKAS.

References

[1] Safinah Ali, Hae Won Park, and Cynthia Breazeal. 2020. Can children emulate a robotic non-player character's figural creativity?. In *CHI PLAY 2020 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 499–509. doi:10.1145/3410404.3414251

[2] Antonio Andriella, Henrique Siqueira, Di Fu, Sven Magg, Pablo Barros, Stefan Wermter, Carme Torras, and Guillem Alenyà. 2021. Do I have a personality? Endowing care robots with context-dependent personality traits. *International Journal of Social Robotics* 13, 8 (2021), 2081–2102. doi:10.1007/s12369-020-00690-5

[3] Sean Andrist, Bilge Mutlu, and Adriana Tapus. 2015. Look like me: Matching robot personality via gaze to increase motivation. In *Conference on Human Factors in Computing Systems - Proceedings*, Vol. 2015-April. 3603–3612. doi:10.1145/2702123.2702592

[4] Julio Cesár Bahamón and R. Michael Young. 2017. An empirical evaluation of a generative method for the expression of personality traits through action choice. In *Proceedings of the 13th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE 2017*. 144–150. doi:10.1609/aiide.v13i1.12951

[5] Murray R. Barrick, Michael K. Mount, and Timothy A. Judge. 2001. Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment* 9, 1-2 (2001), 9–30. doi:10.1111/1468-2389.00160

[6] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. 71-81 pages. Issue 1. doi:10.1007/s12369-008-0001-3

[7] Jasmin Bernotat and Friederike Eyssel. 2017. A robot at home - How affect, technology commitment, and personality traits influence user experience in an intelligent robotics apartment. In *RO-MAN 2017 - 26th IEEE International Symposium on Robot and Human Interactive Communication*, Vol. 2017-Janua. 641–646. doi:10.1109/ROMAN.2017.8172370

[8] Cynthia Breazeal and Rodney Brooks. 2012. Robot emotion. In *Who Needs Emotions?: The Brain Meets the Robot*. Oxford University Press, 271–310. doi:10.1093/acprof:oso/9780195166194.003.0010

[9] Álvaro Castro-González, José Carlos Castillo, Fernando Alonso-Martín, Olmer V. Olortegui-Ortega, Victor González-Pacheco, María Malfaz, and Miguel A. Salichs. 2016. The effects of an impolite vs. A polite robot playing rock-paper-scissors. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9979 LNAI. 306–316. doi:10.1007/978-3-319-47437-3_30

[10] Raymond B Cattell. 1950. *Personality: A systematic theoretical and factual study, 1st ed.* McGraw-Hill, New York, NY, US. vii, 689–vii, 689 pages. doi:10.1037/10773-000

[11] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. Vicuna: An open-source chatbot impressing GPT-4 with 90%* ChatGPT quality. https://lmsys.org/blog/2023-03-30-vicuna/

[12] J. Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

[13] Paul T. Costa and Robert R. McCrae. 2008. The revised NEO personality inventory (NEO-PI-R). In *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing*. SAGE Publications Ltd, 1 Oliver's

Yard, 55 City Road, London EC1Y 1SP United Kingdom, 179–198. doi:10.4135/9781849200479.n9

[14] Henriette Cramer, Jorrit Goddijn, Bob Wielinga, and Vanessa Evers. 2010. Effects of (in)accurate empathy and situational valence on attitudes towards robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 141–142. doi:10.1109/hri.2010.5453224

[15] Kerstin Dautenhahn. 2007. Socially intelligent robots: Dimensions of human-robot interaction. In *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 362. 679–704. doi:10.1098/rstb.2006.2004

[16] Colin G. DeYoung, Lena C. Quilty, and Jordan B. Peterson. 2007. Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology* 93, 5 (2007), 880–896. doi:10.1037/0022-3514.93.5.880

[17] Melissa Donnermann, Philipp Schaper, and Birgit Lugrin. 2022. Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education. *Frontiers in Robotics and AI* 9, March (2022), 1–12. doi:10.3389/frobt.2022.831633

[18] James E. Driskell, Gerald F. Goodwin, Eduardo Salas, and Patrick Gavan O'Shea. 2006. What makes a good team player? Personality and team effectiveness. In *Group Dynamics*, Vol. 10. 249–271. doi:10.1037/1089-2699.10.4.249

[19] Albert Ellis and Mike Abrams. 2009. *Personality theories: Critical perspectives*. Sage Publications, Inc, Thousand Oaks, CA, US. xix, 696–xix, 696 pages.

[20] Connor Esterwood, Kyle Essenmacher, and Han Yang. 2021. A meta-analysis of human personality and robot acceptance in human-robot interaction. In *Conference on Human Factors in Computing Systems - Proceedings*. doi:10.1145/3411764.3445542

[21] Juan Fasola and Maja J. Matarić. 2012. Using socially assistive human-robot interaction to motivate physical exercise for older adults. In *Proceedings of the IEEE*, Vol. 100. IEEE, 2512–2526. doi:10.1109/JPROC.2012.2200539

[22] Alessio Galatolo, Iolanda Leite, and Katie Winkle. 2023. Personality-adapted language generation for social robots. In *IEEE International Workshop on Robot and Human Communication, RO-MAN*. 1800–1807. doi:10.1109/RO-MAN57019.2023.10309335

[23] Kanishk Gandhi, J. Philipp Fränken, Tobias Gerstenberg, and Noah D. Goodman. 2023. Understanding social reasoning in language models with language models. In *Advances in Neural Information Processing Systems*, Vol. 36. arXiv:2306.15448

[24] Lewis R Goldberg. 1990. An alternative "description of personality": The Big-Five factor structure. 1216–1229 pages. doi:10.1037/0022-3514.59.6.1216

[25] Samuel D. Gosling, Peter J. Rentfrow, and William B. Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528. doi:10.1016/S0092-6566(03)00046-1

[26] Barbara J. Grosz. 1996. Collaborative systems. In *AI Magazine*, Vol. 17. 67–85. doi:10.2307/j.ctv36xvqqx.6

[27] David Ha and Douglas Eck. 2017. A neural representation of sketch drawings. *CoRR* abs/1704.03477 (2017). arXiv:1704.03477 http://arxiv.org/abs/1704.03477

[28] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2010. Assessing acceptance of assistive social agent technology by older adults: The Almere model. *International Journal of Social Robotics* 2, 4 (2010), 361–375. doi:10.1007/s12369-010-0068-5

[29] Airlie Hilliard, Cristian Munoz, Zekun Wu, and Adriano Soares Koshiyama. 2024. Eliciting personality traits in large language models. (2024). arXiv:2402.08341 http://arxiv.org/abs/2402.08341

[30] Marius Hoggenmueller, Maria Luce Lupetti, Willem Van Der Maden, and Kazjon Grace. 2023. Creative AI for HRI design explorations. In *ACM/IEEE International Conference on Human-Robot Interaction*. 40–50. doi:10.1145/3568294.3580035

[31] Chien Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. 2015. Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology* 6, July (2015), 1–12. doi:10.3389/fpsyg.2015.01049

[32] Jihong Hwang, Taezoon Park, and Wonil Hwang. 2013. The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. *Applied Ergonomics* 44, 3 (2013), 459–471. doi:10.1016/j.apergo.2012.10.010

[33] Chipp Jansen and Elizabeth Sklar. 2021. Exploring co-creative drawing workflows. *Frontiers in Robotics and AI* 8, May (2021), 1–26. doi:10.3389/frobt.2021.577770

[34] Hang Jiang, Xiajie Zhang, Xubo Cao, Cynthia Breazeal, Deb Roy, and Jad Kabbara. 2024. PersonaLLM: Investigating the ability of Large Language Models to express personality traits. In *Findings of the Association for Computational Linguistics: NAACL 2024*, Kevin Duh, Helena Gomez, and Steven Bethard (Eds.). Association for Computational Linguistics, Mexico City, Mexico, 3605–3627. doi:10.18653/v1/2024.findings-naacl.229

[35] Belhassein Kathleen, Fernández Castro Víctor, Mayima Amandine, Clodic Aurélie, Pacherie Elisabeth, Guidetti Michèle, Alami Rachid, and Cochet Hélène. 2022. Addressing joint action challenges in HRI: Insights from psychology and philosophy. doi:10.1016/j.actpsy.2021.103476

[36] Matthias Kerzel, Erik Strahl, Sven Magg, Nicolás Navarro-Guerrero, Stefan Heinrich, and Stefan Wermter. 2017. NICO—Neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 113–120.

[37] Sara Kiesler and Jennifer Goetz. 2002. Mental models of robotic assistants. In *Conference on Human Factors in Computing Systems - Proceedings*. 576–577. doi:10.1145/506443.506491

[38] Rinaldo Kühne and Jochen Peter. 2023. Anthropomorphism in human–robot interactions: a multidimensional conceptualization. *Communication Theory* 33, 1 (2023), 42–52. doi:10.1093/ct/qtac020

[39] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P. Losey, and Dorsa Sadigh. 2020. When humans aren't optimal: Robots that collaborate with risk-aware humans. In *ACM/IEEE International Conference on Human-Robot Interaction*. 43–52. doi:10.1145/3319502.3374832 arXiv:2001.04377

[40] Minae Kwon, Malte F. Jung, and Ross A. Knepper. 2016. Human expectations of social robots. *ACM/IEEE International Conference on Human-Robot Interaction* 2016-April (2016), 463–464. doi:10.1109/HRI.2016.7451807

[41] Iolanda Leite, André Pereira, Carlos Martinho, and Ana Paiva. 2008. Are emotional robots more fun to play with?. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*. 77–82. doi:10.1109/ROMAN.2008.4600646

[42] Michael Lewis, Katia Sycara, and Phillip Walker. 2018. The role of trust in human-robot interaction. In *Studies in Systems, Decision and Control*. Vol. 117. 135–159. doi:10.1007/978-3-319-64816-3_8

[43] Rosemarijn Looije, Mark A. Neerincx, and Fokie Cnossen. 2010. Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human Computer Studies* 68, 6 (2010), 386–397. doi:10.1016/j.ijhcs.2009.08.007

[44] Todd Lubart, Dario Esposito, Alla Gubenko, and Claude Houssemand. 2021. Creativity in humans, robots, humbots. *Creativity* 8, 1 (2021), 23–37. doi:10.2478/ctra-2021-0003

[45] Cyril Marx, Clemens Könczöl, Agnes Altmanninger, and Bettina Kubicek. 2024. The critical robot: Impact of performance feedback on intrinsic motivation, self-esteem and psychophysiology in human–robot interaction. *International Journal of Social Robotics* 16, 8 (2024), 1751–1764. doi:10.1007/s12369-024-01147-9

[46] Robert R. McCrae and Oliver P. John. 1992. An introduction to the five-factor model and its applications. *Journal of Personality* 60, 2 (jun 1992), 175–215. doi:10.1111/j.1467-6494.1992.tb00970.x

[47] Bernt Meerbeek, Jettie Hoonhout, Peter Bingley, and Jacques M.B. Terken. 2008. The influence of robot personality on perceived and preferred level of user control. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 9, 2 (2008), 204–229. doi:10.1075/is.9.2.04mee

[48] Bernt Meerbeek, Martin Saerbeck, and Christoph Bartneck. 2009. Iterative design process for robots with personality. In *Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*. 94–101. doi:10.13140/RG.2.2.22087.91043

[49] Athanasios Mouratidis, Maarten Vansteenkiste, Willy Lens, and Georgios Sideridis. 2008. The motivating role of positive feedback in sport and physical education: Evidence for a motivational model. *Journal of Sport and Exercise Psychology* 30, 2 (2008), 240–258. doi:10.1123/jsep.30.2.240

[50] George A. Neuman and Julie Wright. 1999. Team effectiveness: Beyond skills and cognitive ability. *Journal of Applied Psychology* 84, 3 (1999), 376–389. doi:10.1037/0021-9010.84.3.376

[51] Tatsuya Nomura and Kazuma Saeki. 2009. Effects of polite behaviors expressed by robots: A case study in Japan. In *Proceedings - 2009 IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2009*, Vol. 2. 108–114. doi:10.1109/WI-IAT.2009.138

[52] Maike Paetzel-Prüsmann, Giulia Perugia, and Ginevra Castellano. 2021. The influence of robot personality on the development of uncanny feelings. *Computers in Human Behavior* 120, September 2020 (2021), 106756. doi:10.1016/j.chb.2021.106756

[53] Miranda A.G. Peeters, Harrie F.J.M. van Tuijl, Christel G. Rutte, and Isabelle Reymen. 2006. Personality and team performance: A meta analysis. *European journal of personality* 20, 5 (2006), 377–396. doi:10.1002/per.588

[54] Miranda A.G. Peeters, Harrie F.J.M. Van Tuijl, Christel G. Rutte, and Isabelle M.M.J. Reymen. 2006. Personality and team performance: A meta-analysis. *European Journal of Personality* 20, 5 (2006), 377–396. doi:10.1002/per.588

[55] Daniel J. Rea, Sebastian Schneider, and Takayuki Kanda. 2021. "Is this all you can do? Harder!": The effects of (im)polite robot encouragement on exercise effort. In *ACM/IEEE International Conference on Human-Robot Interaction*. 225–233. doi:10.1145/3434073.3444660

[56] Jinny Rhee, David Parent, and Anuradha Basu. 2013. The influence of personality and ability on undergraduate teamwork and team performance. *SpringerPlus* 2, 1 (2013), 1–14. doi:10.1186/2193-1801-2-16

[57] Lionel P. Robert Jr., Rasha Alahmad, Connor Esterwood, Sangmi Kim, Sangseok You, and Qiaoning Zhang. 2020. A review of personality in human–robot interactions. *Foundations and Trends® in Information Systems* 4, 2 (2020), 107–212. doi:10.1561/2900000018 arXiv:2001.11777

[58] Richard M. Ryan and Edward L. Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* 25, 1 (2000), 54–67. doi:10.1006/ceps.1999.1020

[59] Richard M. Ryan, Valerie Mims, and Richard Koestner. 1983. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test

using cognitive evaluation theory. *Journal of Personality and Social Psychology* 45, 4 (1983), 736–750. doi:10.1037//0022-3514.45.4.736

[60] Maha Salem, Micheline Ziadee, and Majd Sakr. 2013. Effects of politeness and interaction context on perception and experience of HRI. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 8239 LNAI. 531–541. doi:10.1007/978-3-319-02675-6_53

[61] Maha Salem, Micheline Ziadee, and Majd Sakr. 2014. Marhaba, how may I help you? Effects of politeness and culture on robot acceptance and anthropomorphization. In *ACM/IEEE International Conference on Human-Robot Interaction*. 74–81. doi:10.1145/2559636.2559683

[62] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair!! An interaction with a cheating robot. In *5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*. 219–226. doi:10.1145/1734454.1734546

[63] Ishika Singh, Valts Blukis, Arsalan Mousavian, Ankit Goyal, Danfei Xu, Jonathan Tremblay, Dieter Fox, Jesse Thomason, and Animesh Garg. 2023. ProgPrompt: Program generation for situated robot task planning using large language models. *Autonomous Robots* 47, 8 (2023), 999–1012. doi:10.1007/s10514-023-10135-3

[64] David Sobrín-Hidalgo, Miguel A. González-Santamarta, Ángel M. Guerrero-Higueras, Francisco J. Rodríguez-Lera, and Vicente Matellán-Olivera. 2024. Explaining Autonomy: Enhancing Human-Robot Interaction through Explanation Generation with Large Language Models. (Feb. 2024). http://arxiv.org/abs/2402.04206

[65] Aleksandra Sorokovikova, Natalia Fedorova, Sharwin Rezagholi, and Ivan P. Yamshchikov. 2024. LLMs simulate Big Five personality traits: Further evidence. In *PERSONALIZE 2024 - 1st Workshop on Personalization of Generative AI Systems, Proceedings of the Workshop*. 83–87.

[66] Dag Sverre Syrdal, Kerstin Dautenhahn, Sarah N. Woods, Michael L. Walters, and Kheng Lee Koay. 2007. Looking good? Appearance preferences and robot personality inferences at zero acquaintance. In *Multidisciplinary collaboration for socially assistive robotics - Papers from the 2007 AAAI Spring Symposium, Technical Report (AAAI Spring Symposium - Technical Report)*. 86–92. 2007 AAAI Spring Symposium ; Conference date: 26-03-2007 Through 28-03-2007.

[67] Leila Takayama and Caroline Pantofaru. 2009. Influences on proxemic behaviors in human-robot interaction. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*. IEEE, 5495–5502. doi:10.1109/IROS.2009.5354145

[68] T. L. Tang and Linda Sarsfield-Baldwin. 1991. The effects of self-esteem, task label, and performance feedback on task liking and intrinsic motivation. *Journal of Social Psychology* 131, 4 (1991), 567–572. doi:10.1080/00224545.1991.9713886

[69] Benedict Tay, Younbo Jung, and Taezoon Park. 2014. When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior* 38 (2014), 75–84. doi:10.1016/j.chb.2014.05.014

[70] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and efficient foundation language models. arXiv:2302.13971 [cs.CL]

[71] Jacqueline Urakami and Katie Seaborn. 2023. Nonverbal cues in human-robot interaction: A communication studies perspective. *ACM Transactions on Human-Robot Interaction* 12, 2 (2023). doi:10.1145/3570169

[72] Sarah Theres Völkel and Lale Kaya. 2021. Examining User Preference for Agreeableness in Chatbots. In *ACM International Conference Proceeding Series*. doi:10.1145/3469595.3469633

[73] Michael L. Walters, Dag S. Syrdal, Kerstin Dautenhahn, René te Boekhorst, and Kheng Lee Koay. 2008. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots* 24, 2 (2008), 159–178. doi:10.1007/s10514-007-9058-3

[74] Chao Wang, Stephan Hasler, Daniel Tanneberg, Felix Ocker, Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, and Michael Gienger. 2024. LaMI: Large language models for multi-modal human-robot interaction. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA*, Vol. 1. Association for Computing Machinery, 1–10. doi:10.1145/3613905.3651029 arXiv:2401.15174

[75] Ning Wang, W. Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. 2008. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human Computer Studies* 66, 2 (2008), 98–112. doi:10.1016/j.ijhcs.2007.09.003

[76] Shensheng Wang, Scott O. Lilienfeld, and Philippe Rochat. 2015. The uncanny valley: Existence and explanations. *Review of General Psychology* 19, 4 (2015), 393–407. doi:10.1037/gpr0000056

[77] Zining Wang, Paul Reisert, Eric Nichols, and Randy Gomez. 2024. Ain't misbehavin' - Using LLMs to generate expressive robot behavior in conversations with the tabletop robot Haru. In *ACM/IEEE International Conference on Human-Robot Interaction*. 1105–1109. doi:10.1145/3610978.3640562 arXiv:2402.11571

[78] Andrew B. Williams and Elise Russell. 2014. Humanoid robots discovering creative creative concepts through social interaction. In *AAAI Fall Symposium - Technical Report*, Vol. FS-14-01. 159–160.

[79] Shu Yang, Shenzhe Zhu, Ruoxuan Bao, Liang Liu, Yu Cheng, Lijie Hu, Mengdi Li, and Di Wang. 2024. What makes your model a low-empathy or warmth person: Exploring the origins of personality in LLMs. (2024), 1–25. arXiv:2410.10863 http://arxiv.org/abs/2410.10863

[80] Bowen Zhang and Harold Soh. 2023. Large language models as zero-shot human models for human-robot interaction. In *IEEE International Conference on Intelligent Robots and Systems*. 7961–7968. doi:10.1109/IROS55552.2023.10341488 arXiv:2303.03548

[81] Ceng Zhang, Junxin Chen, Jiatong Li, Yanhong Peng, and Zebing Mao. 2023. Large language models for human–robot interaction: A review. *Biomimetic Intelligence and Robotics* 3, 4 (2023), 100131. doi:10.1016/j.birob.2023.100131

[82] Jakub A. Złotowski, Hidenobu Sumioka, Shuichi Nishio, Dylan F. Glas, Christoph Bartneck, and Hiroshi Ishiguro. 2015. Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology* 6, June (2015), 1–13. doi:10.3389/fpsyg.2015.00883

A Chatbot perception questionnaire

- How relevant and appropriate were the responses of the chatbot? <*Not relevant at all* | *Highly relevant*>
- Did you feel the chatbot responded empathetically? <*Did not empathize at all* | *Fully empathized*>
- How intelligent did you perceive the chatbot to be? <*Not at all intelligent* | *Very intelligent*>
- How easy did you perceive the interaction with the chatbot to be? <*Very difficult* | *Very easy*>
- How likely would you like to have this chatbot as a teammate? <*Very unlikely* | *Very Likely*>
- Did you feel the chatbot understood the queries? <*Did not understand at all* | *Fully understood*>