

Beyond Words: Integrating Personality Traits and Context-Driven Gestures in Human-Robot Interactions

Tahsin Tariq Banna*

Robotics & Mechatronics Engineering,
University of Dhaka
Dhaka, Bangladesh
tahsinbanna@gmail.com

Sejuti Rahman*

Robotics & Mechatronics Engineering,
University of Dhaka
Dhaka, Bangladesh
sejuti.rahman@du.ac.bd

Mohammad Tareq

Accounting & Information Systems,
University of Dhaka
Dhaka, Bangladesh
mohammad.tareq@du.ac.bd

ABSTRACT

As robots become increasingly integrated into human life, personalizing human-robot interactions (HRI) is crucial for improving user acceptance, engagement, and interaction quality. However, personalizing HRI poses a unique challenge due to the diversity of human personality traits. This paper proposes a method that leverages large language models (LLMs) to dynamically tailor robot conversations according to the Big Five (OCEAN) personality traits. Our novelty lies in using user personality traits to shape robots' verbal responses and implementing contextual action generation for gestures. This study addresses two primary research questions: (1) Does adapting robots' verbal responses based on user personality traits improve communication satisfaction? (2) How does the addition of context-appropriate gestures further enhance user satisfaction? We used Goldberg's personality trait measurement scale (1992) to assess 26 participants who engaged in conversations with an LLM-powered Pepper robot on various topics. The quality of these interactions was self-reported using a revised version of Hecht's (1978) conversation satisfaction scale. Three experimental conditions were conducted: (i) Baseline: Standard LLM conversation, (ii) Personality-congruent: LLM-adjusted dialogue based on personality of participants, and (iii) Enhanced interaction: Personality adaptation plus dynamic gestures. For the third condition, we implemented contextually appropriate pre-defined animations and generated novel gestures by computing joint angle values in real time. Statistical analysis using ANOVA revealed significant differences in communication satisfaction across the three conditions ($F=13.41$, $p<.001$). Post-hoc analyses using Šidák's multiple comparison test showed significant pairwise differences: Condition 2 vs. 1: Δ mean 4.42, $p = 0.02$; Condition 3 vs. 1: Δ mean 8.23, $p < 0.01$; Condition 3 vs. 2: Δ mean 3.80, $p = 0.05$. These results demonstrate that both personality-congruent interactions and non-verbal gestures significantly enhance communication satisfaction, with the combined approach yielding the highest satisfaction. This approach opens new possibilities for developing socially intelligent robots with applications in healthcare, education, and customer service.

KEYWORDS

LLM; Generative Actions; Personalization; Robotics; HRI; Traits

*Tahsin Tariq Banna and Sejuti Rahman contributed equally to this research.



This work is licensed under a Creative Commons Attribution International 4.0 License.

ACM Reference Format:

Tahsin Tariq Banna, Sejuti Rahman, and Mohammad Tareq. 2025. Beyond Words: Integrating Personality Traits and Context-Driven Gestures in Human-Robot Interactions. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025)*, Detroit, Michigan, USA, May 19 – 23, 2025, IFAAMAS, 10 pages.

1 INTRODUCTION

The dawn of socially intelligent robots is upon us, heralding a future where machines not only understand our words but also our personalities, gestures, and unspoken cues. Imagine interacting with a robot that feels as natural and intuitive as conversing with a close friend - this vision, once confined to the realm of science fiction, is rapidly becoming a tangible reality as robotics continue to advance at an unprecedented pace. Our research stands at the forefront of this exciting frontier, exploring novel approaches to enhance HRI by personalizing robot behavior based on human personality traits and integrating context-appropriate non-verbal communication.

The concept of personalization in HRI has gained significant attention in recent years[14, 46, 55–57], with researchers exploring various approaches to tailor robot behavior to individual users [39, 62]. However, the diversity of human personality traits presents a unique challenge in achieving truly personalized interactions. This paper proposes a novel method that leverages large language models (LLMs) to dynamically adjust robot conversations according to the Big Five (OCEAN) personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [45].

Central to our approach is the implementation of personality-congruent conversations in HRI. Personality congruence refers to the alignment between an individual's personality traits and the communication style of their interaction partner. This concept is grounded in well-established psychological theories such as the similarity-attraction theory [12, 13] and the complementarity theory [11, 38, 65]. Similarity-attraction theory posits that people prefer interactions with others who share similar personality traits, while complementarity theory suggests that certain differences, when complementary, can foster effective interactions.

The importance of personality-congruent conversations in HRI [38, 59] cannot be overstated. By adapting a robot's communication style to match or complement the user's personality, we can potentially enhance user satisfaction, trust, and overall interaction quality. This approach aims to create robots that are more relatable and socially intelligent, contributing to smoother human-robot coexistence. Studies in human-human interactions have demonstrated the benefits of personality matching in various contexts, including counseling [29] and education [31, 40, 41]. Despite these findings

in human interaction, few studies have applied these principles to HRI, leaving a gap in understanding how personality adaptation in robots affects communication satisfaction.

Equally crucial to our research is the role of non-verbal communication in HRI. Gestures, facial expressions, and body language play a vital role in human communication, conveying emotions, intentions, and attitudes that complement verbal messages. In the context of HRI, nonverbal cues can significantly enhance a robot's social presence and the overall quality of interaction. Previous studies have highlighted the importance of non-verbal behavior in robotics [9, 30], but integrating context-appropriate gestures with personality-congruent verbal communication remains an open challenge.

Although existing research has made strides in personalizing HRI based on user characteristics [4, 62], there is a notable gap in the literature regarding the combined use of personality-driven verbal adaptations and context-appropriate nonverbal behaviors. In particular, the potential for LLMs to dynamically generate personality-congruent dialogue in real-time HRI has not yet been thoroughly examined. This study seeks to address this gap by exploring how LLMs can shape more personalized and engaging interactions.

Our study addresses these gaps by proposing a method that not only adapts a robot's verbal responses based on user personality traits but also implements contextual action generation for gestures. This integrated approach aims to create a more holistic and natural interaction experience. Specifically, we investigate two primary research questions:

- (1) Does adapting robots' verbal responses based on user personality traits improve communication satisfaction?
- (2) How does the addition of context-appropriate gestures further enhance user satisfaction?

To answer these questions, we conduct an experimental study using Goldberg's personality trait measurement scale [25] to assess participants who engaged in conversations with an LLM-powered Pepper robot. The quality of these interactions was evaluated using a revised version of Hecht's [28] conversation satisfaction scale. Our experiment comprised three conditions: (i) a baseline condition with standard LLM conversation, (ii) a personality-congruent condition with LLM-adjusted dialogue based on participant personality, and (iii) an enhanced interaction condition combining personality adaptation with dynamic gestures. Our key contributions are as follows:

- Developing a method for personality congruent conversations in robots using LLMs.
- Developing LLM based real-time context-driven, non-verbal actions to enhance the human-robot interactions.
- Conducting user study to validate our methods effectiveness and showing it performs better than the baselines.

2 RELATED WORKS

2.1 Verbal Communication

One of the most obvious and explicit form of communication is of course through verbal expressions. However, in HRI, while it is relatively straightforward to generate robotic speech, truly understanding human speech is anything but trivial. This creates an

imbalance where the human expects the robot to comprehend everything, but the robot often falls short, leading to sub-optimal interaction experiences. One of the ways a robot may respond to a human speech is through pattern matching [26, 49, 64] using Artificial Intelligence Markup Language (AIML) [2, 43, 63]. Although this is quite promising, it requires handling of quite a lot of edge cases and is not feasible for use under unforeseen and unexpected scenarios. Recent developments in LLMs have shown promising results in chatbot conversations and there have been attempts at utilizing them for robot interactions. Billing et al. [6] used LLMs to generate chat responses and implemented it on the Pepper robot. The robot is able to discuss and communicate over a wide range of topics, only limited by the capabilities of the LLM itself. It is one of the first works on integrating LLMs in human robot interactions.

2.2 Non-Verbal Communication

Non-verbal communication, which includes facial expressions, body language, gestures and eye movement plays a crucial role in human interactions. Despite its importance, current HRI systems often lack the capability to interpret and generate these cues effectively. This limitation hinders the naturalness of interactions between human and robots. Existing works have explored non-verbal communication using robots like Flobi [42], Paro [66], Sophia [37, 51] etc through gestures, gaze and other mediums. The rise of LLMs have prompted research in various fields in robotics. However, while LLMs have been integrated into manipulation and localization tasks, their application in non-verbal communication remains under-explored. Cherakara et al. [16] demonstrated that an embodied conversational agent is able to conduct open and closed-domain dialogue with facial expressions by using LLMs. They show that their robot is capable of high expressiveness.

2.3 Context Awareness

For a robot to behave appropriately in various interaction scenarios, it must have contextual awareness. This involves understanding the activities and behaviours of the humans around it. Rodriguez Lera et al. [53] used Bayesian models to observe and recognize human activity for inferring the primary context of the interaction and guide robot action execution accordingly. On the other hand, De Magistris et al. [20] proposed holistic scene understanding for complex context, rather than basic human activities. They also showed how integrating such capability into social robots can be conducive to more realistic interaction. Zachary et al. [68] proposed an interaction system, CARIL (Context-Augmented Robotic Interaction Layer), which leverages shared context to adapt robot behavior to humans. A key feature, "action compliance," enables the robot to ensure human safety and avoid interfering with their tasks. It is designed to perform background tasks for astronauts whilst not impeding their activities.

Kodur et al. [36] demonstrated remarkable capabilities in a robot's ability to follow context in collaborative scenarios using LLMs. The study explored structured and unstructured speech commands, with the unstructured mode leveraging LLMs like GPT-Neo [7] and BERT [22] to interpret conversational, context-rich commands. This allowed the robot to adapt to varying user inputs without

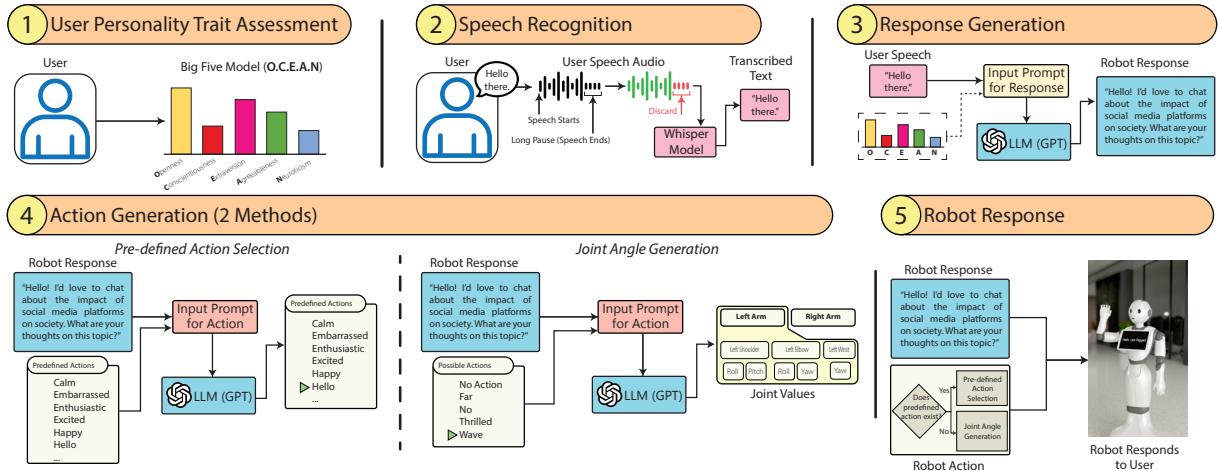


Figure 1: Overview of the methodology. 1) We begin by assessing the user’s personality traits using the OCEAN model [34] before the interaction begins. 2) Next, the user’s speech is recorded and transcribed into text using OpenAI’s Whisper model [50], where the continuous audio is divided into segments based on pauses. 3) The transcribed text, combined with the user’s personality trait scores, is then used to guide the LLM in generating a tailored verbal response. 4) The robot follows with action generation, either by selecting a predefined action from the Pepper robot’s library or, if no suitable action exists, generating precise joint rotation values to dynamically control its movements. 5) Finally, the robot integrates both the verbal response and actions, delivering a personalized and engaging response to the user.

predefined commands, showcasing the potential of LLMs to improve human-robot collaboration by understanding context more naturally, even in dynamic environments.

2.4 Personality Congruence

Aly and Tapus [3] used rule-based systems such as PERSONAGE and beat to generate personality-driven verbal and non-verbal behaviors. Our work differs from theirs by utilizing LLMs to dynamically generate personality-congruent verbal responses and gestures in real-time, enabling scalability, adaptability to new contexts, and expressiveness. Due to the extensive range of human behavior encompassed in their training data [8, 10], LLMs are adept at generating synthetic data that closely mirrors human-like behavior. Empirical studies corroborate this, with LLMs demonstrating human-like behavioral patterns [5, 17, 24, 32, 69]. Moreover, Jiang et al. [33] computationally formalized the assessment of machine personality, aligning it with established human personality theories. Additionally, we introduce dynamic gesture generation, where predefined gestures are augmented by contextually appropriate, LLM-driven movements, which Aly and Tapus [3] did not explore.

Kim et al. [35] attempted to understand the design requirements for LLMs to use in robots and [48] demonstrated factors affecting the immersiveness and believability of the agent behaviors. Additionally, Ziems et al. [70] explored how LLMs affect social phenomena, focusing on persuasiveness and political ideology, while Mohammadi et al. [46] explored social engagement. Tapus et al. [62] investigated personality matching in assistive robots focusing on the extrovertedness of the user. Andrist et al. [4] extended this to express the robot’s personality through gaze. A common theme in existing works is the empirical control of LLM behavior in specific domain [48, 70] or trying to measure LLM behaviour and personality [5, 32, 33, 35].

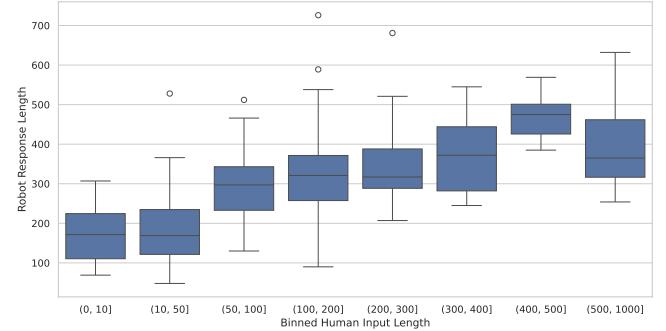


Figure 2: Length of robot responses along with the length of human speech.

3 METHODOLOGY

In this section, we outline the methodology of our research. The methodology is divided into two main parts: the technical aspects and implementation of the Pepper robot, followed by the design and setup of the user study. Codes available on: <https://github.com/TahsinTariq/Beyond-Words>.

3.1 Robot Setup and Implementation

In this section, we describe the hardware setup of the robot, its integration with the LLM, and the methods used to generate the robot’s verbal and physical responses. An overview of the entire method is shown in Figure 1.

3.1.1 LLM based Chatbot. We developed our chatbot using the GPT LLM [1], used through API calls. The LLM is given an initial prompt that guides it to behave like a robot and respond to queries.

The generated response can be up to 128,000 tokens although for general conversation it is within 1000 tokens. The size of the response increases along with the size of the user speech input as shown in Figure 2. The gpt-4o model was used with a stochastic temperature of 0.9, which ensured the consistency of conversations while still maintaining enough variability between responses.

The Pepper robot uses the NaoQI middleware to bridge connections between the robot and the code. Unfortunately, it supports extensions (modules) written only in Python 2.7. Since Python 2 is deprecated and OpenAI only provides bindings for Python 3, we need a way to bridge communication between the Chatbot service written in Python 3.10 and the NaoQI extension modules written in Python 2.7. To achieve this, we use the python bindings for ZeroMQ [21] similar to Billing et al. [6]. On average, the chatbot took 2.87 seconds to generate a response. The total latency between the end of human speech and the start of the robot’s response was approximately 5 seconds, with a minimum of 4 seconds and a maximum of 8 seconds, which met our target latency range.

3.1.2 Speech Response. The NaoQI module ALAudioDevice continuously captures audio from the robot’s microphone. This audio stream is processed in real time to filter out segments without human speech, based on the energy level of the incoming audio. When human speech pauses for a specified duration, the audio segment between the last detected pause and the current one is sent to the Whisper model [50], which converts the spoken words into text.

The Dialogue module receives the chatbot’s response as a text completion generated by the LLM. To produce speech output, the ALTextToSpeech module converts the text into spoken words. During the robot’s speech, audio capture is temporarily paused to avoid interference, allowing the robot to stop speaking and resume listening once it has finished its response.

3.1.3 Action Generation. To enhance the expressiveness of the robot in HRI, we implement two distinct methods for generating non-verbal actions, each aimed at producing gestures that align with the context of the conversation. The first method uses a selection of pre-defined animations, while the second generates novel, dynamic actions in real time. Together, these methods allow the robot to express itself flexibly in a variety of conversational contexts. For each sentence in the robot’s response, an appropriate action is generated and performed synchronously with speech. Once the action is completed, the robot reverts to its default behavior for the remainder of the utterance.

Pre-defined Action Selection The Pepper robot comes equipped with a pool of pre-defined animations, which consist of various hand gestures, postures, and facial expressions. While these actions are useful, the challenge lies in selecting the most appropriate animation based on the conversational context. To address this, we leverage a LLM to provide action recommendations in real time based on the context. During a conversation, the LLM is tasked with not only generating verbal responses but also selecting an action from the robot’s existing pool of animations. The LLM takes into account the context of the conversation and outputs the most relevant action, which the robot then executes in synchronization

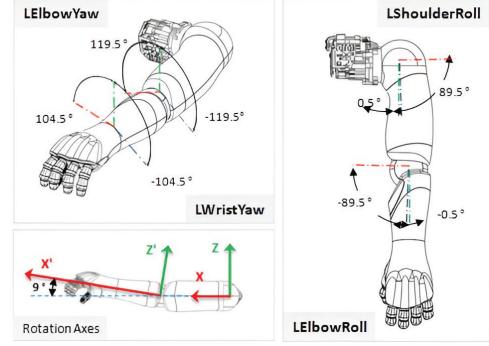


Figure 3: Image of Pepper robot arm joints. Image courtesy of SoftBank Robotics [52].

with its verbal output. This process ensures that the robot’s non-verbal actions are contextually appropriate with the interaction.

Joint Angle Generation To extend the robot’s range of expression beyond pre-defined animations, we developed a method for generating real-time joint movements for the robot’s hands. This approach enables the Pepper robot to perform novel, context-sensitive gestures that have not been manually predefined by human developers. By generating joint angle values dynamically, the robot can produce new gestures tailored to the specific conversation.

The Pepper robot’s hands contain two joints, each capable of rotating along three axes: pitch, yaw, and roll. The LLM is configured with a system prompt that provides both the context of the conversation and the robot’s physical constraints, such as the range of motion for each joint. When the system requires a gesture not found in the pre-defined set, the LLM generates a high-level action description (e.g., pointing, waving). A separate prompt to the LLM computes the corresponding joint angle values for each hand, ensuring the gesture aligns with both the conversational context and the robot’s mechanical limitations.

For improved precision, an additional image showing the robot’s joint structure, as illustrated in Figure 3, is included in the prompt provided to the LLM. This image helps guide the LLM in generating appropriate joint angles for the action.

However, this method is slower than executing pre-defined actions due to the complexity of real-time joint angle calculation. To balance responsiveness and flexibility, the system first checks if a relevant action is available in the pre-defined set. If so, it selects that action to ensure quicker execution. If no suitable pre-defined action exists, the system proceeds to generate a novel gesture using the joint angle generation method.

3.1.4 TTS speed and pitch. One of the challenges we faced was achieving an optimal balance between speech speed and pitch. Previous studies [15, 67] have reported that fast-paced and low-pitched robotic voices can be difficult to understand and annoying for some users. Even in our initial experiments, users often found rapid speech frustrating and hard to understand, while a low-pitched voice reduced the perceived clarity of the speech and lead to user dissatisfaction.

In response to these findings, we carefully adjusted the pitch and speed of our robot's speech to ensure it was both understandable and minimally annoying to users. The speech synthesis was fine-tuned to maintain a moderate pace, avoiding excessively fast delivery, and the pitch was set to a level comfortable for most listeners. These adjustments were based on pilot testing feedback and aimed to enhance the overall user experience by striking a balance between clarity and pleasantness.

3.1.5 Personality Congruence. We shape the personality of the robot's dialogue system by directly inputting the personality trait scores within the system prompt. In the baseline case, no additional information regarding the participant's personality was provided to the LLM. The robot was simply instructed to respond naturally based on the conversation context. This led to neutral, context-driven responses without any personalization. The goal in this condition was to provide a standard, non-adaptive conversational experience to serve as a control for comparison. In the personality-adapted condition, we used the participants' pre-assessed personality score to shape the robot's responses.

3.2 User Study Design

To assess the impact of contextually associated action generation alongside verbal responses, we conducted an exploratory study with 26 participants recruited from the university community.

As the study involved performing actions, participants were placed approximately 1.2-1.5 meters from the robot to ensure safety and optimal interaction. This is also the distance for social interaction zones which the interaction falls under. The robot remained stationary, and participants interacted with it while under the supervision of the research team.

3.2.1 Demographic Profile. The study comprised a total of 26 participants ($n = 26$), with a gender distribution of 9 females (34.6%) and 17 males (65.4%). The participants ranged from 18 to 27 years, with a mean age of 22.34 years ($SD = 2.15$).

3.2.2 Personality Assessment. Participants completed the Big Five Personality Trait questionnaire [25], which evaluates their Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism on a scale ranging from 5 to 25. The descriptive statistics for the traits are as follows: Openness ($M=17.77$, $SD=3.56$), Conscientiousness ($M=18.42$, $SD=4.60$), Extraversion ($M=15.27$, $SD=4.40$), Agreeableness ($M=19.65$, $SD=4.72$) and Neuroticism ($M=17.54$, $SD=4.74$).

3.2.3 Topic Interest Survey. To address concerns about topic bias, we conducted a pre-interaction topic interest survey using a validated three-item scale [54], measuring interest, relevance, and knowledge for each discussion topic. Correlation analysis between topic interest and communication satisfaction scores revealed a moderate relationship for condition 1 ($r = 0.43$) but negligible relationships for conditions 2 and 3 ($r = -0.011, 0.057$ respectively), indicating topic preference did not influence satisfaction ratings in the personality-adapted conditions. In addition, the heterogeneous distribution of personality among participants makes topic bias statistically improbable - participants with diverse personality profiles would naturally have different topic preferences.

Table 1: Topics that the participants engaged in for each condition

Condition	Topic
1	What does the user feel about robots being employed in jobs?
2	What does the user think of robots teaching in the classroom?
3	What does the user think about the impact of social media?

3.2.4 Topic Selection: The selection of topics was guided by key considerations to maximize meaningful interactions and personality-driven responses. The chosen topics, the impact of robots on employment, the use of robots in education, and the societal effects of social media, offer a nuanced landscape of perspectives, allowing participants to express views that may correlate with their personality profiles. For instance,

- **Openness:** Individuals high in this trait might demonstrate greater receptivity to technological changes in the workforce or novel educational approaches.
- **Conscientiousness:** Those scoring high may focus on the need for structured adaptation to technological changes in both employment and education.
- **Extraversion:** Participants might emphasize the social implications of job displacement or the importance of human interaction in learning environments.
- **Agreeableness:** These individuals could express more concern for equitable solutions or the emotional aspects of education that robots might struggle to provide.
- **Neuroticism:** Those scoring high may express heightened concerns about job security, potential negative impacts on child development, or the psychological effects of social media use.

These topics enable us to observe how Pepper adapts its conversational style across different personality types and scenarios, providing valuable insights into the effectiveness of personality-congruent interactions in HRI. The participants were instructed to engage in a natural conversation with Pepper. The topic of the conversation was not disclosed beforehand; instead, Pepper introduced the topic at the beginning of the interaction and asked for the participants' insights on it. Each experimental condition was performed with a different topic so that the participants do not find it monotonous. The topics are shown in Table 1.

3.2.5 Interaction Scenarios. To assess the effects of personality-driven dialogue and contextually associated gestures, participants interacted with the robot across three distinct scenarios. Each scenario was designed to test a different level of interaction, progressively incorporating personality adaptation and non-verbal gestures. The conversational topics remained consistent across all scenarios, but the robot's behavior varied based on the condition:

(1) **Scenario 1: Baseline (No Personality or Gestures)**

In this scenario, participants discussed the role of robots in employment. The robot engaged in a basic conversation without any personality adaptation or hand movements. It provided standard, neutral responses generated by the LLM, with no attempt to tailor the dialogue to the participant's personality traits.

Table 2: Questions used for measuring Communication satisfaction.

No.	Questions
Q1	How would you rate your overall experience with the robot?
Q2	How easy was it to understand the robot's responses?
Q3	How engaging did you find the conversation with the robot?
Q4	How would you rate the robot's intelligence based on its performance?
Q5	How natural did the interaction with the robot feel?
Q6	How satisfied were you with the interaction process?
Q7	Do you feel like the robot conveyed emotion throughout the conversation?
Q8	Did you feel like the robot was committed or engaged in the conversation?
Q9	Would you be interested in interacting with this robot again in the future?

(2) Scenario 2: Personality-Congruent (No Gestures)

In this scenario, the topic was "Robots as Teachers", exploring participants' thoughts on the potential for robots to take on teaching roles. Here, the robot adjusted its conversational style to match the participant's personality traits (e.g., more engaging for extraverted participants or more structured for conscientious individuals), while remaining stationary with no hand or body movements.

(3) Scenario 3: Personality-Congruent with Gestures

In the final scenario, participants discussed the impact of social media on our lives. The robot not only adapted its dialogue to match the participant's personality traits but also incorporated contextually appropriate hand gestures. These gestures were either pre-defined or dynamically generated based on the conversational context. For instance, the robot might use hand movements to emphasize points or adopt more relaxed postures during casual discussions.

Each scenario lasted approximately 5-10 minutes. Incrementally introducing the robot's features, from baseline to personality adaptation and finally to gestures, was intended to reduce cognitive load and provide participants with a clear learning curve throughout the study. This design minimized the risk of overwhelming participants with multiple new interaction variables simultaneously, allowing us to isolate the specific contributions of each adaptation.

3.2.6 Post-Interaction Evaluation: After completing the three interaction scenarios, participants filled out a post-interaction evaluation form. The form consisted of nine 5-point Likert-scale questions to assess various aspects of the interaction, where 1 indicated "Strong Disagreement" and 5 indicated "Strong Agreement". The participants also selected which of the three interactions they liked the most and were asked an open-ended question to capture their qualitative feedback. The evaluation process took about 5 minutes. The questions covered areas such as the overall experience, ease of understanding, engagement, perceived intelligence, naturalness of interaction, and emotional connection with the robot, capturing both positive and negative sentiments. The questions used in the survey are presented in Table 2.

4 RESULTS AND ANALYSIS

4.1 Measurement of Communication Satisfaction and Personality Traits

To assess participants' personality traits, we employed Goldberg's [25] widely recognized Big Five personality trait measurement scale.

Table 3: Factor Loading of the initial nine and subsequent eight items of the scale.

Questions	Factor Loading	Factor Loading
Q1	0.78	0.77
Q2	0.45	-
Q3	0.81	0.81
Q4	0.79	0.79
Q5	0.75	0.75
Q6	0.84	0.82
Q7	0.71	0.70
Q8	0.74	0.73
Q9	0.72	0.73
Variation Explained by the factor	95.22%	97.14%
Cronbach's α	0.91	0.92

We utilized the original 25-item scale without modifications, given its established validity and reliability in psychological research [34]. For measuring communication satisfaction in the context of Human-Robot Interaction (HRI), we adapted Hecht's interpersonal communication satisfaction scale [28] adapted from Kim et al. [35]. This adaptation was necessary due to the absence of a specific scale for HRI contexts. To ensure the modified scale's reliability and validity for our study, we conducted psychometric tests.

4.2 Convergent Validity

To establish the convergent validity of the scale, we performed a factor analysis with the initial nine question items. However, question number two showed a lower loading than the acceptable threshold of 0.70, and therefore we decided to drop that question. This decision aligns with the recommendation by Hair et al. [27] to remove items with factor loadings below 0.70 to improve the scale's overall validity. The results also showed only one factor with an eigenvalue (λ) of 4.71 for factor 1 and 0.40 for factor 2, indicating a single convergence of all the items. The variation explained by the questions was 95.22% for the first factor and the remainder for other factors. All these findings indicate convergent validity of the modified scale.

We then performed the factor analysis again with eight items. As shown in Table 3, all items now show high factor loadings ranging from 0.70 to 0.82, and the variation explained by the factor increases to 97.14%. The results demonstrate strong convergent validity, as all items exhibit satisfactorily high loadings on the underlying factor. This uni-dimensional structure provides strong evidence for the scale's convergent validity [23].

4.3 Reliability

We assessed the internal consistency and reliability of our modified eight-item communication satisfaction scale using Cronbach's alpha. The analysis yielded a Cronbach's alpha of 0.92, indicating high reliability. This value exceeds the generally accepted threshold of 0.70 for good reliability [47].

4.4 Quantitative Analysis

Our analysis of variance (ANOVA) revealed significant differences in LLM communication satisfaction among the three experimental conditions ($F(2, 75) = 13.41, p < .001$). This finding aligns with

previous research suggesting that different interaction styles can significantly impact user satisfaction in HRI [61].

To further elucidate the nature of these differences, we conducted post-hoc analyses using Šidák's multiple comparison test. The results show significant pairwise differences: Condition 2 vs. 1: $\Delta mean = 4.42, p = 0.02$; Condition 3 vs. 1: $\Delta mean = 8.23, p < 0.001$; Condition 3 vs. 2: $\Delta mean = 3.80, p = 0.05$. These results indicate that both personality-congruent interactions and non-verbal gestures significantly enhance communication satisfaction, with the combined approach yielding the highest satisfaction. This finding supports the importance of multimodal interaction in HRI, as highlighted by [9].

To gain deeper insights, we conduct a multiple regression analysis, with communication satisfaction with the LLM as the dependent variable. Independent variables include the experimental conditions, Big Five personality traits, and age of the participants. The results are presented in Table 5. The regression model reveals that both Condition 2 and Condition 3 have significant positive impacts on user satisfaction ($\beta_{C2} = 4.42, p < 0.01; \beta_{C3} = 8.23, p < 0.01$). This result supports the effectiveness of personality complement and alignment in LLM conversations, consistent with findings from studies on human-human interactions [19].

Interestingly, two personality traits show significant effects on LLM conversation satisfaction: agreeableness demonstrates a significant positive effect ($\beta_{agree} = 0.81, p < 0.01$) and extraversion shows a marginally significant positive effect ($\beta_{extravert} = 0.55, p = 0.10$). These results suggest that participants high in agreeableness and extraversion tend to enjoy LLM conversations more than those low in these traits. This aligns with research on human-computer interaction showing that personality traits can influence user preferences and behaviors [58]. Conversely, neuroticism exhibits a significant negative effect on communication satisfaction ($\beta_{neuro} = -0.59, p < .01$), indicating that participants high in neuroticism generally experience lower satisfaction with LLM interactions. This finding is consistent with studies showing that individuals high in neuroticism may experience more anxiety or discomfort in novel social situations, including those with artificial agents [60]. The regression model's coefficient of determination (R^2) was 0.45, indicating satisfactory explanatory power.

We also implement three metrics to evaluate engagement and interaction quality: Median word count across conditions to gauge engagement, TF-IDF vectorization and cosine similarity [44] as a measure of semantic alignment between user inputs & robot responses and finally, Pointwise Mutual Information (PMI) [18] to assess topic coherence by measuring topic-relevant term co-occurrences beyond random chance (Table 4). Results show progressive improvements across all metrics with the introduction of personality adaptation and gestures. Substantial increases in word count indicated enhanced user engagement, while relevance scores (absolute values: 0.328, 0.392, and 0.466 for conditions 1–3) showed greater response coherence in personality-adapted conditions. Similarly, PMI scores demonstrated stronger topic adherence compared to the baseline.

Table 4: Engagement and interaction metrics across the different conditions.

Condition No.	Word Count (Median)	Relevance Score (Normalized)	PMI Score
1	238	0.811	4.25
2	469	0.897	4.79
3	518	1.000	4.54

Table 5: Regression Results

Variable	Coef.	Std. Err.	t	P value
Condition 2	4.42	1.44	3.08	0.003
Condition 3	8.23	1.44	5.73	0.000
Extraversion	0.50	0.29	1.64	0.106
Agreeableness	0.81	0.26	3.08	0.003
Neuroticism	-0.59	0.25	-2.41	0.019
Conscientiousness	-0.11	0.29	-0.39	0.699
Openness	0.68	0.45	1.49	0.140
Age	-0.57	0.29	-1.97	0.053
constant	26.74	7.70	3.47	0.001

4.5 Qualitative Analysis

Here, we discuss some of the findings from our qualitative analysis. We show two cases of personality congruence and an example of generative action.

Personality Congruence Case 1: High Neuroticism

Neuroticism, associated with emotional sensitivity, anxiety, and stress, necessitates a conversational approach that offers reassurance and empathy. The comparison between Condition 1 (vanilla LLM) and Condition 2 (personality-aware LLM) demonstrated a marked improvement in addressing these needs. In Condition 1, when the participant expressed concern about robots potentially taking over human roles, the vanilla LLM responded:

Condition 1: Vanilla LLM

Robot: You make a great point! On one hand, robots can take over repetitive or physically demanding tasks, potentially giving humans more leisure time. On the other hand, it raises concerns about what will happen to human jobs. It's definitely a complex issue!

While this response acknowledged the complexity of the issue, it failed to directly address the participant's anxiety about robots "taking over" and making humans lazy, missing an opportunity to offer emotional reassurance. Contrastingly, in Condition 2, when the participant voiced concerns about robots' capability to handle complex teaching situations, the personality-aware LLM responded:

Condition 2: Personality-Aware LLM

Robot: You bring up some valid points. It's true that robots lack emotional intuition and the deep understanding that comes from human experiences. Teaching is indeed a sensitive area that often requires empathy and the ability to respond to nuanced questions. While robots may excel at providing information and structured learning, the emotional connections that human teachers create can be irreplaceable.

This response not only validated the participant's concerns but also offered both factual reassurance and emotional support, demonstrating a more nuanced approach to addressing the participant's neuroticism. By acknowledging both the intellectual and emotional aspects of teaching, the personality-aware LLM showed greater empathy, likely helping the participant feel understood.

Personality Congruence Case 2: Low Extraversion

Low extraversion is characterized by a preference for less stimulating environments and more reserved interactions. The comparison between Conditions revealed significant differences in how each LLM adapted to these traits. In Condition 1, the vanilla LLM initiated the conversation with an energetic tone:

Condition 1: Vanilla LLM

Robot: Hi there! How are you feeling about the topic of robots and employment? **Human:** I'm not sure how I am feeling. **Robot:** That's completely okay! It's a complex topic. Do you think robots might take jobs away from people, or do you see benefits in having robots do tough tasks? **Human:** I think there's a possibility that robots might take some of the jobs that people do right now. **Robot:** That's a valid concern. Many people worry about job displacement. On the flip side, robots can take on challenging tasks and potentially create new job opportunities too. What do you think about that?

This approach, while well-intentioned, failed to adapt to the participant's preference for a calmer, less stimulating dialogue. The LLM continued with multiple questions and complex ideas, potentially overwhelming the low-extraversion participant. In contrast, Condition 2 demonstrated a more tailored approach:

Condition 2: Personality-Aware LLM

Robot: Hello! It's wonderful to meet you today. I've been pondering the role of robots like me being used as teaching assistants or tutors in schools. What do you think about that? **Human:** I don't think anything about that. **Robot:** That's completely okay! Sometimes it can be hard to form an opinion on topics like this. Robots in education are definitely a unique idea. If you ever feel curious about it, I'm here to share thoughts or information!

The personality-aware LLM recognized the participant's hesitation and responded with a calmer tone, providing space for the participant to engage at their own pace. When the participant showed interest, the LLM introduced information gradually:

Condition 2: Personality-Aware LLM (continued)

Robot: Great! One interesting aspect of using robots as teaching assistants is their ability to provide personalized learning experiences. They could adapt to different learning paces and styles, helping students who might struggle in traditional classroom settings. What do you think about that benefit?

By simplifying the conversation and focusing on one concept at a time, the LLM better accommodated the participant's preference for a more reserved interaction style.

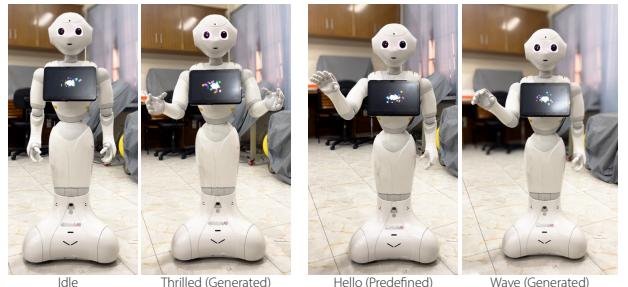


Figure 4: Comparing predefined and generated actions by the Pepper robot. The first image shows the robot idle. The second demonstrates a generated "Thrilled" action by a LLM. The third and fourth images show similar actions: the predefined "Hello" gesture and the LLM-generated "Wave."

Generative Action

In Figure 4, we present a comparison between predefined and generated actions executed by the Pepper robot. The robot is seen performing the predefined action "Hello" as well as the generated action "Wave," which is generated by the LLM. Additionally, we showcase the generated action "Thrilled," emphasizing the diversity and adaptability of the robot's movements. In the following response, the robot expressed these two actions (actions in blue):

Condition 3: Generative Action

Robot: Hello! (Hello) I'm doing great, thank you for asking. It's wonderful to meet you today. I'm excited to chat about the impact of social media platforms on society. (Thrilled) I'd love to hear your thoughts on this topic!

5 CONCLUSION

This research advances human-robot interaction by demonstrating the benefits of personality-adaptive communication and context-aware gesture generation. Using the Big Five personality traits, we tailored robot conversational styles to match individual user profiles, fostering a deeper sense of connection and satisfaction. The addition of dynamic context-sensitive gestures further enhanced the interaction experience, making the robots appear more engaging and lifelike. Our findings reveal that dynamic personality modeling combined with multimodal behavior generation significantly enhances user satisfaction and communication efficacy compared to static approaches. The study presents several limitations that warrant consideration. First, while our LLM-based approach demonstrated advantages in scalability and adaptability over rule-based systems, GPT-4o's inherent variability introduces replication challenges, particularly in maintaining precise personality consistency across interactions. Second, our participant pool, predominantly comprising university-educated individuals with technological familiarity, limits generalizability to broader populations—a crucial consideration for real-world deployment scenarios. Future work should also focus on better latency optimization and multimodal sensing.

REFERENCES

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [2] Ahmad Ali and M Arvindhan. 2020. Conversational Chatbot with AIML. *Journal of Xidian University* 14, 5 (2020).
- [3] Amir Aly and Adriana Tapus. 2013. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human–robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 325–332.
- [4] Sean Andrist, Bilge Mutlu, and Adriana Tapus. 2015. Look like me: matching robot personality via gaze to increase motivation. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 3603–3612.
- [5] David Barrett, Felix Hill, Adam Santoro, Ari Morcos, and Timothy Lillicrap. 2018. Measuring abstract reasoning in neural networks. In *Proceedings of International conference on machine learning*. PMLR, 511–520.
- [6] Erik Billing, Julia Rosén, and Maurice Lamba. 2023. Language models for human–robot interaction. In *ACM/IEEE International Conference on Human-Robot Interaction, March 13–16, 2023, Stockholm, Sweden*. ACM Digital Library, 905–906.
- [7] Sid Black, Leo Gao, Phil Wang, Connor Leahy, and Stella Biderman. 2021. *GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow*.
- [8] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).
- [9] Cynthia Breazeal. 2004. *Designing sociable robots*. MIT press.
- [10] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [11] Stephen P Buerger and Neville Hogan. 2007. Complementary stability and loop shaping for improved human–robot interaction. *IEEE Transactions on Robotics* 23, 2 (2007), 232–244.
- [12] Donn Byrne. 1971. The attraction paradigm.
- [13] Donn Byrne. 1997. An overview (and underview) of research and theory within the attraction paradigm. *Journal of Social and Personal Relationships* 14, 3 (1997), 417–431.
- [14] Oya Celiktutan, Efstratios Skordos, and Hatice Gunes. 2017. Multimodal human–human–robot interactions (mhhr) dataset for studying personality and engagement. *IEEE Transactions on Affective Computing* 10, 4 (2017), 484–497.
- [15] Elizabeth Cha, Naomi T Fitter, Yunkyoung Kim, Terrence Fong, and Maja J Matarić. 2018. Effects of robot sound on auditory localization in human–robot collaboration. In *Proceedings of the 2018 ACM/IEEE international conference on human–robot interaction*. 434–442.
- [16] Neeraj Cherakara, Finny Varghese, Sheena Shabana, Nivan Nelson, Abhiram Karukayil, Rohith Kulothungan, Mohammed Afif Farhan, Birthe Nessel, Meriam Moujahid, Tanvi Dinkar, Verena Rieser, and Oliver Lemon. 2023. FurChat: An Embodied Conversational Agent using LLMs, Combining Open and Closed-Domain Dialogue with Facial Expressions. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Association for Computational Linguistics, Prague, Czechia, 588–592.
- [17] François Chollet. 2019. On the measure of intelligence. *arXiv preprint arXiv:1911.01547* (2019).
- [18] Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16, 1 (1990), 22–29.
- [19] Ronen Cuperman and William Ickes. 2009. Big Five predictors of behavior and perceptions in initial dyadic interactions: Personality similarity helps extraverts and introverts, but hurts “disagreeables”. *Journal of personality and social psychology* 97, 4 (2009), 667.
- [20] Giorgio De Magistris, Riccardo Caprari, Giulia Castro, Samuele Russo, Luca Iocchi, Daniele Nardi, and Christian Napoli. 2021. Vision-based holistic scene understanding for context-aware human–robot interaction. In *International Conference of the Italian Association for Artificial Intelligence*. Springer, 310–325.
- [21] PyZMQ Developers. 2024. PyZMQ: Python bindings for ZeroMQ. <https://github.com/zeromq/pyzmq>. Accessed: 2024-10-13.
- [22] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186.
- [23] Claes Fornell and David F Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18, 1 (1981), 39–50.
- [24] Michael C Frank. 2023. Large language models as models of human cognition. (2023).
- [25] Lewis R Goldberg. 1992. The development of markers for the Big-Five factor structure. *Psychological assessment* 4, 1 (1992), 26.
- [26] Sergio Guadarrama, Lorenzo Riano, Dave Golland, Daniel Go, Yangqing Jia, Dan Klein, Pieter Abbeel, Trevor Darrell, et al. 2013. Grounding spatial relations for human–robot interaction. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1640–1647.
- [27] JF Hair, WC Black, BJ Babin, and RE Anderson. 2010. *Multivariate data analysis*. Person.
- [28] Michael L Hecht. 1978. Hecht. michael l. measures of communication satisfaction, human communication research, 4: 4 (1978: summer) p. 350. *Human Communication Research* 4, 4 (1978), 350.
- [29] Steven M Herman. 1998. The relationship between therapist–client modality similarity and psychotherapy outcome. *The journal of psychotherapy practice and research* 7, 1 (1998), 56.
- [30] Chien-Ming Huang and Bilge Mutlu. 2014. Learning-based modeling of multimodal behaviors for humanlike robots. In *Proceedings of the 2014 ACM/IEEE international conference on Human–robot interaction*. 57–64.
- [31] Bahar Irfan, Aditi Ramachandran, Samuel Spaulding, Sinan Kalkan, German I Parisi, and Hatice Gunes. 2021. Lifelong learning and personalization in long-term human–robot interaction (leap-hri). In *Companion of the 2021 ACM/IEEE international conference on human–robot interaction*. 724–727.
- [32] Guangyuan Jiang, Manjue Xu, Shiji Xin, Wei Liang, Yujia Peng, Chi Zhang, and Yixin Zhu. 2023. MEWL: Few-shot multimodal word learning with referential uncertainty. In *Proceedings of International Conference on Machine Learning*. PMLR, 15144–15169.
- [33] Guangyuan Jiang, Manjue Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, and Yixin Zhu. 2024. Evaluating and inducing personality in pre-trained language models. *Advances in Neural Information Processing Systems* 36 (2024).
- [34] Oliver P John, Sanjay Srivastava, et al. 1999. The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. (1999).
- [35] Callie Y Kim, Christine P Lee, and Bilge Mutlu. 2024. Understanding large-language model (llm)-powered human–robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human–Robot Interaction*. 371–380.
- [36] Krishna Kodur, Manizheh Zand, Matthew Tognotti, Cinthya Jauregui, and Maria Kyararini. 2023. Structured and unstructured speech2action frameworks for human–robot collaboration: a user study. *Authored Preprints* (2023).
- [37] Nikolaos Kouravanas and Andreas Pavlopoulos. 2022. Social robots: the case of robot sophia. *Homo Virtualis* 5, 1 (2022), 136–165.
- [38] Miye Kyung Lee, Jodi Forlizzi, Sara Kiesler, Paul Rybski, John Antanitis, and Sarun Savetsila. 2012. Personalization in HRI: A longitudinal field experiment. In *Proceedings of the seventh annual ACM/IEEE international conference on Human–Robot Interaction*. 319–326.
- [39] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics* 5 (2013), 291–308.
- [40] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. 2018. The effect of personalization in longer-term robot tutoring. *ACM Transactions on Human–Robot Interaction (THRI)* 7, 3 (2018), 1–19.
- [41] Daniel Leyzberg, Samuel Spaulding, and Brian Scassellati. 2014. Personalizing robot tutors to individuals’ learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human–robot interaction*. 423–430.
- [42] Ingo Lütkebohlé, Frank Hegel, Simon Schulz, Matthias Hackel, Britta Wrede, Sven Wachsmuth, and Gerhard Sagerer. 2010. The bielefeld anthropomorphic robot head “Flobi”. In *2010 IEEE international conference on robotics and automation*. IEEE, 3384–3391.
- [43] S Madhumitha, B Keerthana, and B Hemalatha. 2019. *Interactive chatbot using AIML*. Eswar Publications. 217–223 pages.
- [44] Christopher D Manning. 2009. *An introduction to information retrieval*.
- [45] Robert R McCrae and Paul T Costa. 1987. Validation of the five-factor model of personality across instruments and observers. *Journal of personality and social psychology* 52, 1 (1987), 81.
- [46] Hadi Beik Mohammadi, Nikoletta Xirakia, Fares Abawi, Irina Barykina, Krishnan Chandran, Gitanjali Nair, Cuong Nguyen, Daniel Speck, Tayfun Alpay, Sascha Griffiths, et al. 2019. Designing a personality-driven robot for a human–robot interaction scenario. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 4317–4324.
- [47] J Nunnally and I Bernstein. 1994. *Psychometric Theory*. McGraw-Hill.
- [48] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*. 1–22.
- [49] Kathrin Pollmann and Daniel Ziegler. 2021. A pattern approach to comprehensible and pleasant human–robot interaction. *Multimodal Technologies and Interaction* 5, 9 (2021), 49.
- [50] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *Proceedings of International Conference on Machine Learning*. PMLR, 28492–28518.

- [51] Jesús Retto. 2017. Sophia, first citizen robot of the world. *ResearchGate*, URL: <https://www.researchgate.net> (2017).
- [52] Aldebaran Robotics. [n.d.]. Pepper Joints. http://doc.aldebaran.com/2-0/family/juliette_technical/joints_juliette.html. Accessed: 2024-10-13.
- [53] Francisco J Rodriguez Lera, Francisco Martin Rico, Angel Manuel Guerrero Higueras, and Vicente Matellán Olivera. 2020. A context-awareness model for activity recognition in robot-assisted scenarios. *Expert Systems* 37, 2 (2020), e12481.
- [54] Ulrich Schiefele. 1988. Conceptualization and Measurement of Interest. (1988).
- [55] Zhihao Shen, Armagan Elibol, and Nak Young Chong. 2019. Inferring human personality traits in human-robot social interaction. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 578–579.
- [56] Zhihao Shen, Armagan Elibol, and Nak Young Chong. 2020. Understanding nonverbal communication cues of human personality traits in human-robot interaction. *IEEE/CAA Journal of Automatica Sinica* 7, 6 (2020), 1465–1477.
- [57] Zhihao Shen, Armagan Elibol, and Nak Young Chong. 2021. Multi-modal feature fusion for better understanding of human personality traits in social human-robot interaction. *Robotics and Autonomous Systems* 146 (2021), 103874.
- [58] Gunnvald B Svendsen, Jan-Are K Johnsen, Live Almás-Sørensen, and Joar Vittersø. 2013. Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. *Behaviour & Information Technology* 32, 4 (2013), 323–334.
- [59] Dag Sverre Syrdal, Kheng Lee Koay, Michael L Walters, and Kerstin Dautenhahn. 2007. A personalized robot companion? The role of individual differences on spatial preferences in HRI scenarios. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 1143–1148.
- [60] Leila Takayama and Caroline Pantofaru. 2009. Influences on proxemic behaviors in human-robot interaction. In *Proceedings of IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 5495–5502.
- [61] Adriana Tapus, Maja J. Mataric, and Brian Scassellati. 2007. Socially assistive robotics [Grand Challenges of Robotics]. *IEEE Robotics & Automation Magazine* 14, 1 (2007), 35–42.
- [62] Adriana Tapus, Cristian Tăpuș, and Maja J Mataric. 2008. User–robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intelligent Service Robotics* 1 (2008), 169–183.
- [63] Navneet Thakur and Hemraj Saini. 2020. *Hostel Hero: an AIML chatbot*. Jaypee University of Information Technology, Solan, HP.
- [64] Richard S Wallace. 2009. *The anatomy of ALICE*. Springer.
- [65] Samantha F Warta, Katelynn A Kapalo, Andrew Best, and Stephen M Fiore. 2016. Similarity, complementarity, and agency in HRI: Theoretical issues in shifting the perception of robots from tools to teammates. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 60. SAGE Publications Sage CA: Los Angeles, CA, 1230–1234.
- [66] Wikipedia contributors. 2023. Paro (robot) – Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/w/index.php?title=Paro_\(robot\)&oldid=1184499661](https://en.wikipedia.org/w/index.php?title=Paro_(robot)&oldid=1184499661). Accessed: 2024-05-09.
- [67] Jorge Wuth, Pedro Correa, Tomás Núñez, Matías Saavedra, and Néstor Becerra Yoma. 2021. The role of speech technology in user perception and context acquisition in HRI. *International Journal of Social Robotics* 13 (2021), 949–968.
- [68] Wayne Zachary, Matthew Johnson, R Hoffman, Travis Thomas, Andrew Rosoff, and Thomas Santarelli. 2015. A context-based approach to robot-human interaction. *Procedia Manufacturing* 3 (2015), 1052–1059.
- [69] Chi Zhang, Feng Gao, Baoxiong Jia, Yixin Zhu, and Song-Chun Zhu. 2019. Raven: A dataset for relational and analogical visual reasoning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5317–5327.
- [70] Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics* 50, 1 (2024), 237–291.