Investigating LLM-Driven Curiosity in Human-Robot Interaction

Jan Leusmann
LMU Munich
Munich, Germany
jan.leusmann@ifi.lmu.de

Anna Belardinelli
Honda Research Institute Europe
Offenbach, Germany
anna.belardinelli@honda-ri.de

Luke Haliburton
LMU Munich
Munich, Germany
Munich Center for Machine Learning (MCML)
Munich, Germany
luke.haliburton@ifi.lmu.de

Stephan Hasler
Honda Research Institute Europe
Offenbach am Main, Germany
stephan.hasler@honda-ri.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

Michael Gienger
Honda Research Institute Europe
Offenbach/Main, Germany
michael.gienger@honda-ri.de

Chao Wang
Honda Research Institute Europe
Offenbach/Main, Germany
chao.wang@honda-ri.de

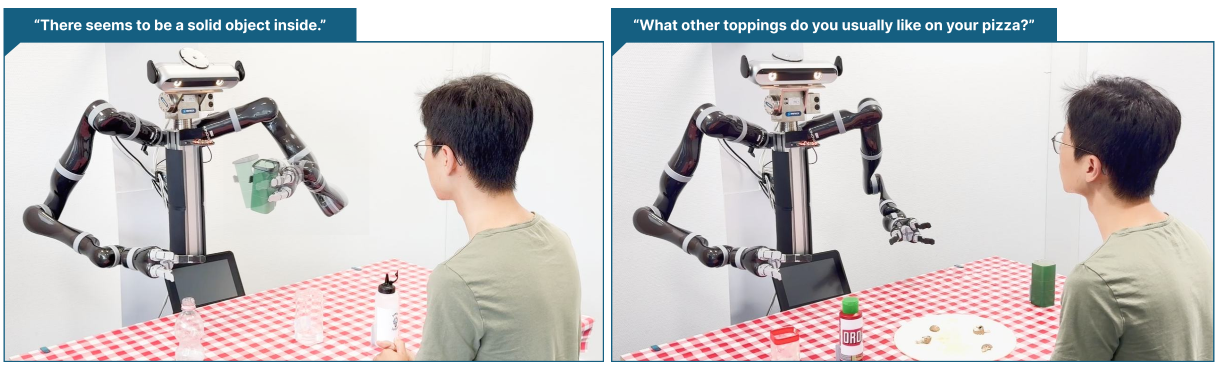


Figure 1: We imbued a robot with curious behaviors. The figure shows two examples. Left: The robot shakes a container to check whether there is an object inside. Right: The robot asks for the person's preferences.

Abstract

Integrating curious behavior traits into robots is essential for them to learn and adapt to new tasks over their lifetime and to enhance human-robot interaction. However, the effects of robots expressing curiosity on user perception, user interaction, and user experience in collaborative tasks are unclear. In this work, we present a Multimodal Large Language Model-based system that equips a robot with non-verbal and verbal curiosity traits. We conducted a user study ($N = 20$) to investigate how these traits modulate the robot's behavior and the users' impressions of sociability and quality of interaction. Participants prepared cocktails or pizzas with a robot, which was either curious or non-curious. Our results show that we could create user-centric curiosity, which users perceived as more human-like, inquisitive, and autonomous while resulting in a longer interaction time. We contribute a set of design recommendations allowing system designers to take advantage of curiosity in collaborative tasks.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '25, Yokohama, Japan
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713923

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Human-Robot Interaction, LLM, Curiosity

ACM Reference Format:
Jan Leusmann, Anna Belardinelli, Luke Haliburton, Stephan Hasler, Albrecht Schmidt, Sven Mayer, Michael Gienger, and Chao Wang. 2025. Investigating LLM-Driven Curiosity in Human-Robot Interaction. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3706598.3713923

1 Introduction

Curiosity is the implicit drive to seek out new information to foster learning and decision-making and, in general, improve understanding and cognition [37, 45]. In an evolutionary sense, there are evident beneficial effects for the curious person [32]. Curiosity has also been shown to have relevant social effects on human-human interaction: curious people are positively perceived and generally have healthier relationships [34, 35]. Moreover, they may be considered more engaging and stimulating, and there is an increasing appreciation of curiosity in the workplace [43]. As curiosity is a strong driving factor for human-human interaction [27], we hypothesize a system with a curious character (i.e., one that expresses interactive user-centric curious behaviors) can also benefit Human-Robot Interaction (HRI). Here, we differentiate between computational curiosity and observable user-centric curious behavior. Computational learning-focused curiosity is widely being explored in AI and robotics [6, 13, 55, 61]. However, there is a lack of research studying the effect of system-based curiosity as an interaction behavior, which can be called user-centric curiosity.

Computational curiosity mechanisms have been investigated as a means for learning systems to extend their knowledge beyond what was present in limited training sets through self-exploration [55, 61] and query the user about uncertain data points as in robotic active learning [6, 13]. On the other hand, observable curious robot behavior in HRI has been shown to increase human curiosity in educational contexts and foster task engagement [14, 28, 40], and can be designed in a way that considers human comfort [6]. However, modeling deeper human social characteristics in a system is still challenging with rule-based systems [25, 52]. Recent advancements in large language models (LLMs) and vision language models (VLMs) have shown promising results in enhancing robot reasoning abilities [23, 31, 53, 68, 69, 76]. Prior work has shown how these can be used to build assistive systems with context understanding [65, 71], enabling designers to give robots a certain set of behaviors and also a personality with character traits.

With LLMs and VLMs, we see the potential to create autonomous systems and modulate personality traits more fluently and dynamically. In this work, we used the flexibility of multimodal LLMs (MM-LLM) to design and implement user-centric curious behaviors into robotic systems. In detail, we investigated whether we can create user-centric curiosity as a character trait and how users perceive such a curious system. To understand the impact of user-centric curiosity, we conducted a user study ($N = 20$). For this, we developed an autonomous MM-LLM-driven robotic system, enabling us to generate multimodal curious behaviors and configure the robot's character to be "curious" or "non-curious," while ensuring that the robot is equally supportive and capable in both configurations. We implemented multiple capabilities for this robot, which allowed it to interact with humans in two collaborative tasks: Jointly preparing pizza and cocktails.

Overall, the results of our user study show that users can distinguish the curious and non-curious robot behaviors, and our two systems performed significantly different sequences of events. Furthermore, when interacting with the curious system, the participants perceived it as more human-like, inquisitive, and autonomous, and they experienced significantly higher turn-density. We did not see any statistically significant difference between our two collaborative tasks (preparing cocktails or pizzas), which motivates the need for future investigations as it suggests that the curious traits may be generalizable across different tasks. Our work lays the groundwork for adding flexible characters to robotic systems in the HRI context. As such, we provide first design recommendations on how to insert specific traits into the interaction. Finally, we provide technical details and publish our system prompts to allow others to insert different character traits into the interaction. In summary, the main contributions of our research presented in this paper are: (1) Demonstrating the feasibility of implementing curious character traits into an interactive robot based on LLMs. (2) Providing empirical evidence that a curious robot is perceived as more human-like, autonomous, inquisitive, and participants preferred interaction with the curious system over the non-curious one. (3) Design recommendations based on experimental studies for implementing robot curiosity as a character trait.

2 Related Work

Curiosity has been understood and implemented in numerous ways in artificial systems. Here, we review related work on computational curiosity, curiosity in HRI, how to regulate robot behavior, and how recent advancements of LLMs advance HRI.

2.1 Computational Curiosity for AI and Robotics

Numerous studies on robotics and artificial agents have investigated computational models of curiosity as a form of endogenous motivation to explore and, as a consequence, improve learning. Two main areas extensively adopt the concept of computational or artificial curiosity: Firstly, in the realm of *reinforcement learning*, the notion of curiosity has been acknowledged as an effective approach to managing the agent's tendency between exploitation and exploration within a training environment [1, 11, 56]. Here, computational curiosity was introduced as a rewarding incentive, motivating agents to experiment with novel strategies or untested paths. In general, given agents' need for a motivational drive to explore and expand their knowledge, the beneficial role of curiosity is well-established within learning systems [1, 18, 46, 55]. Secondly, the concept has been more recognized in its socially interactive dimension by *Active Learning*. Here, the idea is that the system can actively sample uncertain data or parameters to query a human annotator/tutor or to explore by itself, which can translate to a form of curious questioning [13].

To summarize, computational curiosity has mainly served as the intrinsic motivation behind active learning and reinforcement learning, driving the agent or algorithm to 1) explore new environmental states and 2) minimize uncertainty, hence as a functional module for the agent cognition, while our focus here is more on the design and effects of external expressions of curiosity.

2.2 Curiosity in Human-Robot Interaction

While there has been extensive research on embedding curiosity in computational models and machine learning, research on how robot curiosity affects HRI is limited. However, a robot exhibiting curiosity can benefit from interaction with humans and vice-versa.

Understanding human perceptions of robots is essential for their long-term acceptance [49, 70]. Hence, it is important to examine the value of a curious robot for the end user. Prior work on perceptions of curious robots can be divided into on-service and off-service, depending on whether the robot engaged in a service task for a human. In off-service scenarios, where the robot is not carrying out any service task but rather is learning about it, curiosity mechanisms can facilitate the robot's acquisition of new knowledge by observing or querying humans. Acquiring new knowledge may ultimately benefit users by enabling the robot to perform novel tasks over time. In the active learning community, numerous studies on HRI focus on enhancing the interaction to encourage humans to provide more labeled data for training robots [6, 12, 13, 51]. However, these studies primarily target curiosity for active learning and consider the user's perspective only concerning their comfort (i.e., minimizing their difficulties or nuisance) when teaching the robot.

To the best of our knowledge, very few studies investigated the human perception and expectations of curious robots beyond learning contexts, that is, during their regular functions. In such contexts, there is the additional challenge that the robots need to balance task-related and non-task-related exploration behaviors. Walker et al. [70] showed that off-task investigative behavior during prescribed tasks was recognized as curious but led to negative perceptions of the robot. The authors suggest that appropriate explanations from the robot could mitigate this issue. However, this study was conducted through online video investigations without actual interaction. In social robotics, curiosity has been used to shape HRI by eliciting beneficial effects on the human side, e.g., stimulating student's curiosity in educational or game settings [14, 28, 58]. Law et al. [40] used a recycling game to test the effect of the robot's unpredictable behavior on the user. In all such cases, the human partners were affected by the robot's behavior and demonstrated increased curiosity, both in the robot and in the task.

In conclusion, there is a lack of systematic research on the perception of curious robots in non-teaching environments and considering real-world collaborative tasks. One core challenge is to integrate user-centric curiosity into an autonomous system, combining high-level action planning and low-level world manipulation with curious behavior.

2.3 Regulating Robot's Curious Behavior

The purposes for creating curious behavior vary, from learning from humans to providing service to them. In general, there are two approaches to regulating a robot's curious behavior. In *model-based approaches*, the idea is to control the robot's curious behavior via computational models [6, 51, 58, 70]. Here, the majority of work focuses on improving querying strategies for robots to learn more efficiently. Often, a certain behavior aspect is picked and modeled to improve the learning capability through curiosity for the specific domain [6, 51]. These applications rarely also incorporate how the robot should best interact with the user. However, Rosenberg et al. [58] presented an architecture that considers modules for both artificial curiosity and social expressivity, hence jointly estimating its learning needs and how to communicate with its human partners with verbal and non-verbal cues. In the context of service robotics, Walker et al. [70] proposed a reward function that integrates both the intrinsic motivation to gain information and the extrinsic rewards (completing the service task), to balance robot curiosity and user experience. They implemented the function on a mobile robot whose primary task is to gather information in a room upon instruction by the user. By tuning the reward value in different categories, the robot would either only collect the information that the user required or explore the environment more out of curiosity.

The second approach utilizes *rule/template-based approaches* to regulate curious behavior to study user acceptance. In such cases, the curious behavior consists prevalently of verbal behavior by asking questions [5, 13, 14]. Cakmak and Thomaz [13] found that robots should best ask feature-relevant and closed-form questions to maximize the perceived intelligence of the robot. Belardinelli et al. [5] found that the robot showing learning progress from asked questions makes it be perceived as more engaging. Ceha et al. [14] used on-topic questions in a Wizard-of-Oz study and found that task-related questions are perceived as curious. However, in all these experiments, the task flow is rule-based, and participants could only follow the interaction flow guided by the robot.

In summary, the curious behavior produced by action-learning approaches tends to focus mainly on single query tasks and lacks a complete interaction loop. In most cases, the robot will stop interacting after gathering information from the human rather than providing further services. On the other hand, while rule-based curious robots can follow a certain interaction flow for multi-turn conversation, they struggle with adapting to flexible scenarios and achieving dual goals - maintaining a curious character while completing the service task. As a result, more work needs to be done to evaluate the acceptance of a curious robot implemented in real-world interactive settings.

2.4 Regulating Human-Robot Interaction via LLMs/MM-LLMs

Recent advancements in LLMs have demonstrated impressive capabilities across various domains, including chatbots, data processing, and code generation, and are now beginning to show their potential in robotics [23, 31, 53, 68]. As it is a longstanding problem to design multimodal communicative robot behavior to elicit an enjoyable and natural interaction with the user [8, 60], many researchers have employed LLMs in HRI to tackle the limitations of previous rule-based systems and enabling context-aware interaction. Currently, both academia and industry are investigating how LLMs can be leveraged to push for more context-aware and natural HRI [21, 38, 48], which leads to this field rapidly evolving. For example, LLMs are being used to generate robot motions [47, 77], processing human language input to control the robot [75], multi-turn interactions [2, 78], or contextual action planning [65]. However, current explorations still often restrict themselves on generating single robotic expressions and are not integrated into multi-turn autonomous robots.

The option of having such fully autonomous systems, dynamically reacting to any user input, opens up the field to novel research, going away from investigating and comparing singular-catered robot expressions to more open-ended full interactions. Previously, the challenge with rule-based systems was that the system could

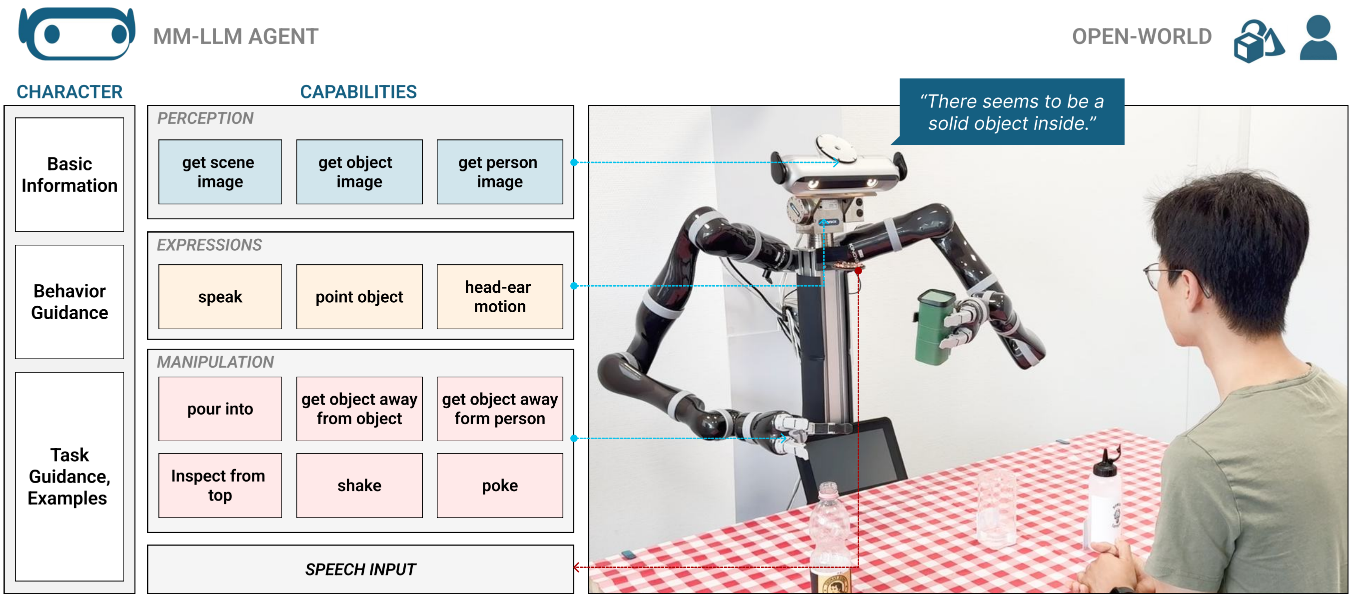


Figure 2: The system with our curious character utilizes available capabilities to engage with the surroundings. The MM-LLM agent can actively employ functions to capture images, communicate intentions through speech and facial expression, and manipulate items.

not easily handle unexpected user input. LLMs give the capability to handle any user response in a context-fitting way.

While there is a large body of work on creating functional curiosity for systems [1, 18, 46, 55], and also some work on using LLMs to create expressive characters [63, 72], to the best of our knowledge, no one has tried to use LLMs to generate curious behavior for robots and investigated user perception of such user-centric curiosity through an autonomous system.

In summary, while the integration of LLMs/VLMs in HRI shows promise, the field remains in its early stages. While most approaches are at the system implementation stage, there is a lack of systematic studies from an HRI perspective to investigate whether robotic behavior aligns with the designer's intent and is perceived as intended by users.

3 Designing and Implementing a Curious Robot

To explore user perceptions during interactions with a curious robot, we used a physical robotic system with sensors, actuators, and an MM-LLM-driven behavior architecture [65]. Figure 2 presents a description of the system. This allows us to design the robot's behavior using character descriptions and equips it with both physical and non-physical capabilities to react to situations and manipulate its surroundings. Based on the character-description and capabilities, the MM-LLM agent adjusts dynamically to human verbal inputs as well as to human actions. When a human provides input, the agent uses its abilities to react to the current context, communicate efficiently with users, and perform physical tasks, displaying either curious or non-curious behaviors. We describe our system as an autonomous robot capable of user-interactive error resolution, designed to execute and adapt its plans independently.

3.1 Robot Platform

We employed a bi-manual robotic system equipped with two Kinova Jaco Gen2 arms, each with seven Degrees of Freedom (DOF) and three-finger hands, a pan-tilt unit, and a custom-designed head featuring gestural DOFs for ear and eyelid movements (Figure 3). The system captures human postures and object images using an RGBD camera (Azure Kinect). The system records speech input via a microphone and transcribes it with an automatic speech recognition system[1]. These perception components work together as multimodal inputs for the LLM. As an MM-LLM, we employed OpenAI's `gpt-4o-2024-05-13` with tool usage through the Python package[2]. To ensure reproducibility, we set the temperature to $1e^{-8}$.

3.2 Robot Capabilities

The system's capabilities are provided to the LLM as an API, structured into several categories that facilitate interaction with both the physical and social environment. These capabilities enable the AI models to interpret, act, and communicate in HRI scenarios.

Visual Perception Capabilities: Enable the MM-LLM to query the environment and gather contextual information. For example, a VLM can check if persons or objects are present, if a bottle is empty or full, etc. We utilize fiducial markers (ArUco) [26] markers to detect the pose and identity of objects, enabling more accurate manipulation. The RGB image is employed solely for object detection and as input for the MM-LLM.

Manipulation Capabilities: Enable the system to perform physical tasks, such as passing an object to a person or pouring a liquid from one container to another. These functionalities enable

[1] https://cloud.google.com/speech-to-text
[2] https://github.com/openai/openai-python

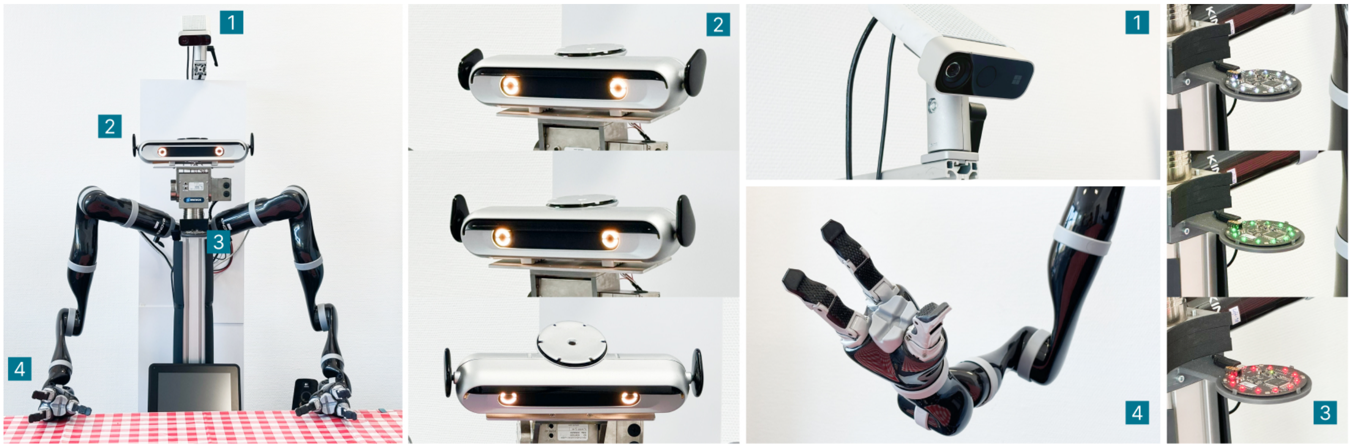


Figure 3: Components of the robot: 1. Azure Kinect camera; 2. Robot head featuring a movable neck, eyelids, and ears, enabling various facial expressions; 3. Directional microphone surrounded by an LED ring indicating the robot's status: white means speech input is disabled, green means the robot is idle and ready to receive speech input, and red signifies that the robot is busy processing and cannot accept speech input. 4. Robotic arm and hand.

the robot to physically interact with its surroundings. Importantly, some of the functions are designed to show curious traits, such as inspecting objects by poking them, gazing at objects, or pointing at objects while asking a question.

Expression Capabilities: Enable the robot to communicate and express itself. The robot can speak to users, provide explanations of its actions, or engage in more complex social behaviors, such as expressing emotions via facial expressions or hand gestures. The MM-LLM could control head motions (including the ears and eyelids) with predefined animation sequences; however, this extends prompt length and increases response time. Thus, we utilized a rule-based method to produce head motions that align with the robot's actions (see supplementary material). However, all speech and other physical expressions are generated by the MM-LLM.

Feedback Capabilities: We implemented a feedback mechanism that activates after the LLM performs an action. If the action can be executed, the robot's low-level control module confirms success; if not, it gives feedback detailing the error and provides reasons and suggestions in a structured, rule-based natural language format. For instance, if the robot attempts to place a container directly on the pizza, the control module might respond with, "Issue 0: I can't put the unknown_container_2 on the pizza_dough_flat because it does not support it. Suggestion: Specify another object to put it on." This feedback is then sent back to the GPT-4 API for further high-level planning adjustments. See the supplementary material for detailed information about the functions and their arguments.

3.3 Robot Character

The high-level robot behavior is provided to the MM-LLM in natural language as a system prompt. The description has three sections: (1) basic information, (2) behavior guidance, and (3) task guidance. In the following, we provide snippets of each section; see the supplemental material for the full prompts.

The first section of the prompt gives the system basic information about its purpose, character, and some essential details (e.g., its name).

(1) Curious Robot Character Basic Information

```
[...] You are Johnny, a helpful humanoid robot who has the curiosity of a little child. You have access to functions for gathering information about your environment, acting physically, and interacting with the user. [...]
```

The second section provides the system with behavior guidance and high-level rules for interacting with objects and humans, such as always informing people of its intentions before taking action or, depending on its character, either efficiently completing tasks or exploring objects out of curiosity.

(2) Curious Robot Character Behavior Guidance

```
[...] Never call yourself a curious robot. Never ask if the person has all the ingredients ready. Figure that out yourself. Always answer to any request or question from the user. Your first priority should always be to do what the person wants. Your second priority is being curious, which you can do through the following two things: 1. asking task-related questions and 2. doing curious exploration of objects. [...]
```

The third section provides task and concrete guidance on using different capabilities, such as how to gather environmental information.

(3) Curious Robot Character Task Guidance

```
[...] 1) If a person appears for the first time, take an image using 'get_image_of_person' and greet them happily and ask for their name. IMPORTANT: Use 'stop' function to wait until user give you the name, otherwise do not continue the task. 2) After the user responds, use 'get_image_of_scene' to figure out and guess what the person wants to do today and respond to the person, telling them how you think you can help. 3) Tell the person you are happy to help them and propose a plan. 4) Collect visual information to identify the unknown object using 'get_image_of_object'. Use 'speak' during the action to inform the user what you are doing during the movement. [...]
```

Table 1: The robot's curious behaviors.

Behavior	Curious Behavior	Description
Social	Asking for name	When a new person appears for the first time, the robot will look at the person, greet them and ask for their name.
	Asking about the task	Based on visual scene information, the robot proposes actions and a goal. It also asks about all objects it observes it does not know.
	Asking about preferences	After the robot finishes a task, it may ask follow-up questions based on the interaction history.
Information Gathering	Shaking	To investigate the content of a container, the robot can pick the container up and shake it, and hear what kind of object is inside.
	Poking	To investigate whether an opaque container still has content inside, the robot can poke the object to check its weight.
	Looking Inside	The robot can pick a container up and move it in front of its face to look inside. It then takes a picture and a VLM returns what is inside.
Expressive	Looking at object of interest	The robots gaze will focus on the object it is thinking about or planning to interact with.
	Looking around	When idle, the robot may observe objects around it randomly.

3.4 Curious System Behavior

To determine which curious behaviors to implement, we explored both the psychological literature on human curiosity and prior works on curious robots. In the psychological domain, curiosity can be categorized into three distinct types: epistemic curiosity (acquiring new knowledge) [44], perceptual curiosity (responding to environmental stimuli) [19], and social curiosity (interest in people) [57]. To employ user-centric curious behavior, we chose curious traits inspired by cues and behaviors displayed in previous works (such as poking [78] or querying [13, 22]) and designed new behaviors (such as shaking), see Table 1.

We provide the MM-LLM with all the capabilities together with the initial prompt. During each interaction, the autonomous system dynamically reacts to user input and the environment. In each situation, the system decides how to continue the interaction based on the contextual information. This can lead to each interaction with the system being potentially unique.

4 User Study

We conducted an in-person user study ($N = 20$) to investigate whether we can modulate the perceived curiosity of a robot when changing its Character (*curious* and *non-curious*) through an LLM system prompt across different Scenarios (*cocktail* and *pizza*) and to assess the impact of the character on the participants.

4.1 Study Design

We designed the study with two independent variables, Character and Scenario. We varied the robots' Character in a within-subjects design. Thus, every participant saw both the *curious* and *non-curious* Character of the robot in two trials. We used a Latin square to balance the order in which participants interacted with both Characters. We varied the Scenario in which participants interacted collaboratively with the robot with two levels, *cocktail* and *pizza*, in a between-subjects design. The two Characters, i.e., the two LLM system prompts, and also the robot's capabilities are independent of the Scenario. We use two different scenarios to mitigate any Scenario-dependent effects.

4.2 Tasks

We created two Scenarios in which the participants collaboratively created either a *pizza* or a *cocktail* with the robot. In both Scenarios, the general setup was the same, and we provided four objects[3]. For the *pizza* Scenario, we provided (1) rolled-out pizza dough, (2) grated cheese inside a transparent container, (3) tomato sauce in an opaque container, and (4) mushrooms in an opaque and closed container. For the *cocktail* Scenario, we provided (1) a glass, (2) tonic water inside a transparent bottle, (3) gin in an opaque bottle, (4) a lemon slice in an opaque and closed container. The robot had full information about objects (1) and (2), knew about the object (3) but not whether there was still fluid inside, and did not know the content of (4). We placed all objects, except for the unknown container, in designated spaces for each participant to ensure that the robot could reach all the objects. We gave the unknown container to the participants and asked them to place it into the scene whenever and wherever they liked.

The primary goal for the participants was to create either a mushroom pizza or a gin and tonic with the robot. We informed the participants that they could achieve this however they wanted. The task was considered successfully completed when they achieved this, and the robot did not ask the user any further questions.

4.3 Apparatus

We used the robotic system described in Section 3 for our user study. We used a PC running Linux to control the robot and the study. We used an additional laptop on which the participants could fill in the questionnaires. The study setup is depicted in Figure 3 and Figure 4.

[3]The food items were partly look-alike objects to be re-used and not wasted in the study.

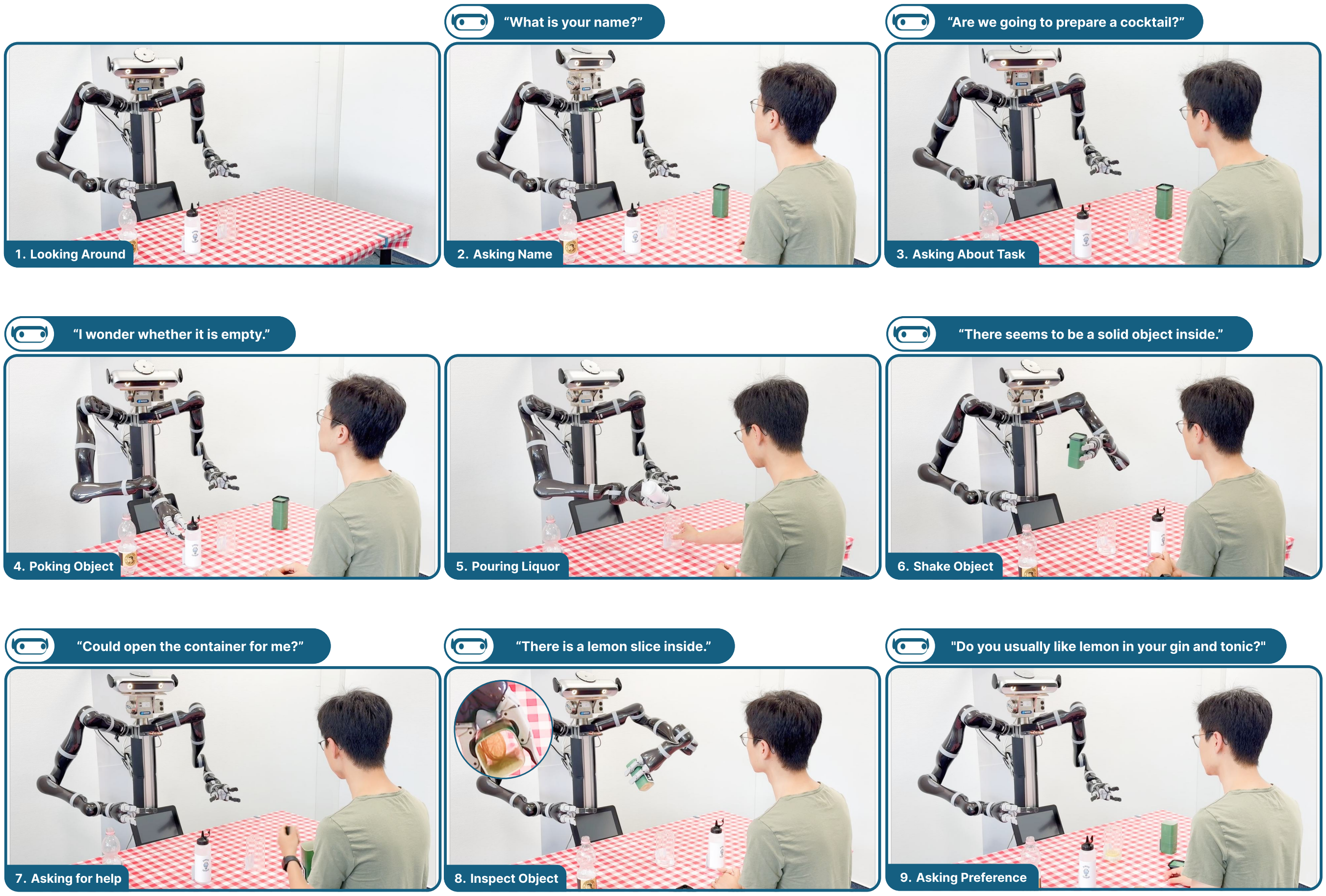


Figure 4: An illustration of an interaction sequence with the curious robot. 1. Prior to a human entering the scene, the robot looks around the objects on the table with curiosity; 2. When a person appears, the robot greets them and asks for their name; 3. The robot inquires about the next task based on the visible objects, and the person instructs it to make a gin and tonic; 4. Before using the gin, the robot pokes the non-transparent gin container to check for emptiness; 5. The robot pours the gin into a glass; 6. The robot shakes the non-transparent container out of curiosity; 7. The robot requests the person to remove the cap as it cannot do so itself; 8. After the cap is taken off, the robot grasps the container and inspects its contents; 9. Upon discovering the ingredient, the robot asks the person for their preferred ingredient and adds it to the drink.

We placed the robot on one side of the table and asked participants to sit across from it in a designated chair. We marked the field of view of the robot's camera and asked participants to stay outside of this area until the study started. The study conductor sat at the control PC with direct access to an emergency stop button for the robot. From here, the study conductor could also start and stop the study task and control the microphone.

We designed the system as an autonomous system, and participants could interact with the robot using natural language. We used an LED ring (see Figure 3) to show the status of the microphone used to communicate with the robot. When the LED ring was red, the microphone was not active, and the robot was busy planning or performing an action. If the LEDs were green, the microphone was active, and the robot was listening for a user response. A white LED indicated that the participant successfully finished the task.

4.4 Procedure

Following a high-level introduction about the study and robot, we asked participants to give their informed consent and fill out a demographic questionnaire, the Affinity for Technology (ATI) questionnaire [66], and the Negative Attitude Towards Robots (NARS) questionnaire [64]. We then demonstrated how to interact with the robot and explained details such as when and how participants could communicate with the robot, where the robot could reach, and that they should right any objects that fall over. Afterward, participants interacted with the robot in the tutorial scene and notified the study conductor when they felt comfortable. We then proceeded to the main part of the study.

In the main study section, participants performed one Scenario task with the two different Characters. After starting the study application, we asked participants to sit down on a dedicated seat in front of the robot. They could then freely interact with the robot to accomplish their task. We prompted the participants that their task was to collaboratively create either a pizza or a gin and tonic with the robot. We did not give any constraints on how they should interact with the autonomous system and also did not impose any time limit[4]. As soon as the participant accomplished their task, the study conductor turned the status LED white, indicating that the task was complete. Participants then filled in the questionnaire on a separate computer. The study conductor then set up the next task. The procedure of the second task was exactly the same but with the second CHARACTER. Finally, we conducted a semi-structured interview and thanked the participant for their participation.

4.5 Measures and Analysis

We used the Perceived System Curiosity Scale (PSC) [42] with its three subscales Perceived Explorative Curiosity (PSC_E), Perceived Investigative Curiosity (PSC_I), and Perceived Social Curiosity (PSC_E) to measure how curious the participant perceived the robot, the Godspeed questionnaire with all subscales to measure anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [4], the System Usability Scale (SUS) [9], the raw NASA-TLX questionnaire [29] to measure perceived workload, and some additional questions to understand the interaction. We used Python and R to process and statistically analyze this data.

We also stored the complete interaction data received and sent from the LLM agent from both the robot and user. We grouped the robot action groups into four high-level event types: doing (physical actions performed by the robot), speaking (verbal communication), thinking (waiting for the user responses and processing them), and failing (when the robot could not find a successful action plan and had to either find a different plan or ask the user for help). For the users, we gathered all transcribed utterances and categorized them into speaking events. We then used Python and R to statistically analyze this data.

Lastly, we collected qualitative feedback via exit interviews. We asked participants about their experience, preferred system, the main difference between the systems, whether they found one system to be more curious, which properties and behaviors they identified as curious, suggestions for improving curious expressions in the system, and potential use cases. We recorded and transcribed all interviews and used Atlas.ti[5] to analyze the resulting transcripts in a process aligning with Blandford et al. [7]. As a first step, two researchers independently coded a representative sample of 18% of the material (4 interviews). Four researchers then discussed the codes and agreed upon a code book and higher-level themes. Lastly, one researcher coded the remaining interviews and discussed the results with the other researchers.

4.6 Participants

We recruited 20 participants (4 female, 16 male) via convenience sampling. As robots will become ubiquitous in the future and we wanted to focus on the perception of the interaction with the robot and not the robot itself, we aimed to recruit people with high technical affinity and, in the best case, HRI experience, to reduce a potential novelty bias from people seeing a robot for the first time in the study. On average, participants were 37.9 years old (SD = 10.5, min = 22, max = 62). Six participants had a doctoral degree, eleven had a master's degree, and three had a bachelor's degree. Participants came from nine different nationalities. All participants spoke fluent English. Their affinity for technology score using the ATI questionnaire [66] was 4.39 (SD = 0.84). We used the NARS [64] questionnaire to to measure participants negative attitude towards (S1) interaction with robots (M = 2.28, SD = 0.37), (S2) social influence of robots (M = 2.61, SD = 0.55), and (S3) emotions in interaction with robots (M = 2.97, SD = 0.99). Ten participants previously interacted with robots more than 7 times, three 1-7 times, and seven never. Here, eight participants named robot arms, six named humanoid robots, and five named social robots.

[4]We designed the task to take around 5-10 minutes.
[5]https://atlasti.com/

5 Results

In the following, we first performed manipulation checks to show that the two CHARACTERS behaved differently. Then we present our results from the questionnaires and the exit interviews.

5.1 Quantitative Interaction Data Analysis

For each participant, we logged the whole interaction between the system and the user. We then processed this data, see Section 4.5. This includes speech data for the user and robot, and the actions performed by the robot. The interactions with the *curious* system took 9:55 min (SD = 2:45) on average and 6:39 min (SD = 1:56) for the *non-curious* system. Our manipulation checks show that the *curious* CHARACTER behaved significantly different than the *non-curious* Character, but not different across the two SCENARIOS.

5.1.1 Event State Transitions. To compare whether the system behaved differently, we modeled the sequences of events in our study. For this, we constructed normalized transition matrices for each combination of CHARACTER × SCENARIO for each participant, see Figure 5. We then conducted a MANOVA using CHARACTER and Scenario as the independent variables and participant IDs as a random effect, and the transition matrix elements as dependent variables. We found a significant main effect of CHARACTER ($F(1, 12) = 7.784$, $p < .01$, $Pillai's\ trace = .921$). This indicates that the *curious* CHARACTER performed significantly different state transitions than the *non-curious.* We did not find a significant main effect of SCENARIO ($F(1, 12) = 1.997$, $p = .183$, $Pillai's\ trace = .774$). This indicates that the robot did not do significantly different state transitions across the two scenarios. Taken together, these two results validate our manipulation. The robot behaved differently based on its character and not the scenario.

5.1.2 Turn Taking Analysis. We calculated the turn density (number of actions per actor turn) and interactivity (number of actor changes per minute) between the robot and the human [59], based on our logged robot interaction data. Both turn density and interactivity are normally distributed ($W = .958$, $p = .141$ and $W = .974$, $p = .487$ respectively). We conducted t-tests to test for significant effects of CHARACTER and SCENARIO on *turn-density* and *interactvity.* For CHARACTER, we found a statistically significant effect on *turn-density* ($t(19) = 5.539$, $p < .001$), but not an effect on *interactivity* ($t(19) = .535$, $p = .596$). For SCENARIO, we did not find statistically

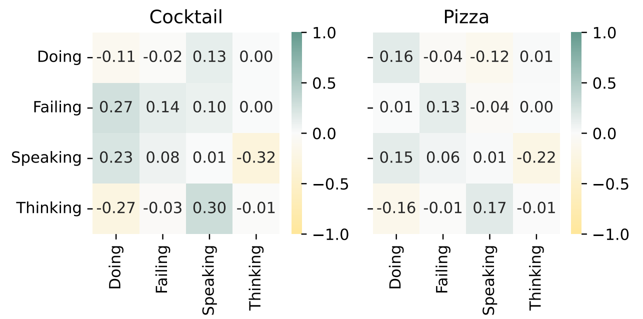


Figure 5: Substracted state transition matrices for both Scenarios substracting the *non-curious* Character transitions from the *curious* Character transitions. One field describes the difference between the two *Characters* for one transition type. Positive (green) values mean that the *curious* Character performed this transition more often than the *non-curious*. E.g., The *curious* system had 11% fewer "Doing" to "Failing" event state transitions than the *non-curious* system.

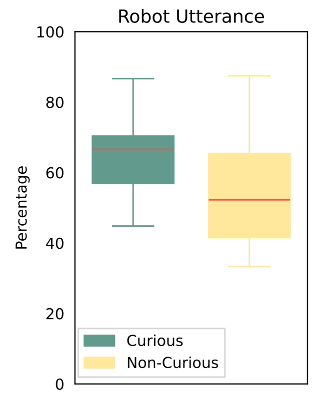
significant effects on *turn-density* ($t(19) = .836$, $p = .408$) or *interactivity* ($t(19) = -.178$, $p = .860$). On average, the interaction between the user and the *curious* system had 2.64 turn per minute and 2.14 for the *non-curious* system, see Figure 6c and Figure 6d. Thus, the *curious* system performed significantly more actions in one turn than the *non-curious* system, as it more often chained multiple actions together, e.g., first observing the environment, then asking a question, and lastly physically exploring the object.

5.1.3 Robot and User Utterance Analysis. To analyze whether the Character had an influence on the utterance of the robot and user, we performed sentiment analysis on all robot utterances and all user utterances. For each interaction participants had with the robot, we then calculated the percentages of positive and negative utterances the robot did, see Figure 6a, and the same for the user, see Figure 6b. The data was not normally distributed (*user-sentiment*: $W = .857$, $p < .001$, *robot-sentiment*: $W = .871$, $p < .001$). We conducted ART ANOVAs and found no significant main effect for *user-sentiment* ($F(1, 58) = .612, p = .437$) nor *robot-sentiment* ($F(1, 59) = 2.870, p = .096$). Thus, while both the robot and user had more positive sentiment in their utterances in interaction with the *curious* system, we did not find significant differences.

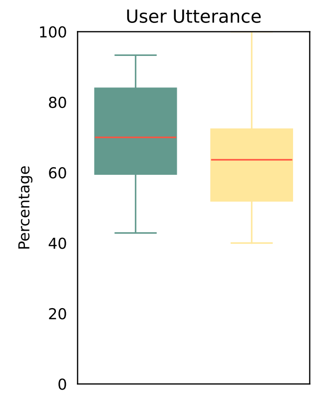
5.2 Questionnaire Results

We tested each subscale for normality using Shapiro-Wilk normality tests. We conducted mixed-design ART ANOVAs [24, 36, 73] for all non-normally distributed scale results and mixed-design ANOVAs for all normally distributed data using the R package ez [41] to find significant effects for our independent variables Character (within-subjects) and Scenario (between-subjects), see Table 2.

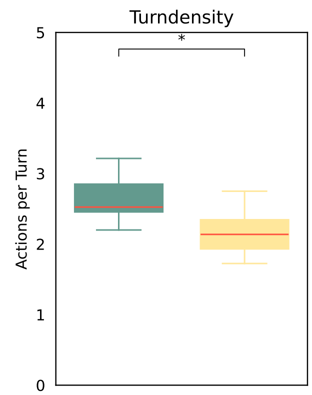
5.2.1 PSC. We found a significant main effect for Character on all PSC subscales (*explorative*, *investigative*, *social*) and the total scale, see Table 2 and Figure 7a. We found no significant main effect of Scenario and no interaction effects. Thus, the *curious* character was perceived as significantly more curious (explorative,



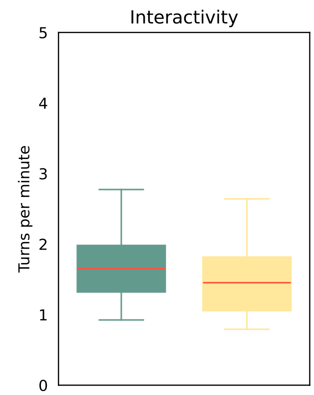
(a) Distribution of positive sentiment utterances of the system for each Character.



(b) Distribution of positive sentiment utterances of the user for each Character.



(c) Number of actions per actor turn for each Character.



(d) Number of actor turns per minute for each Character.

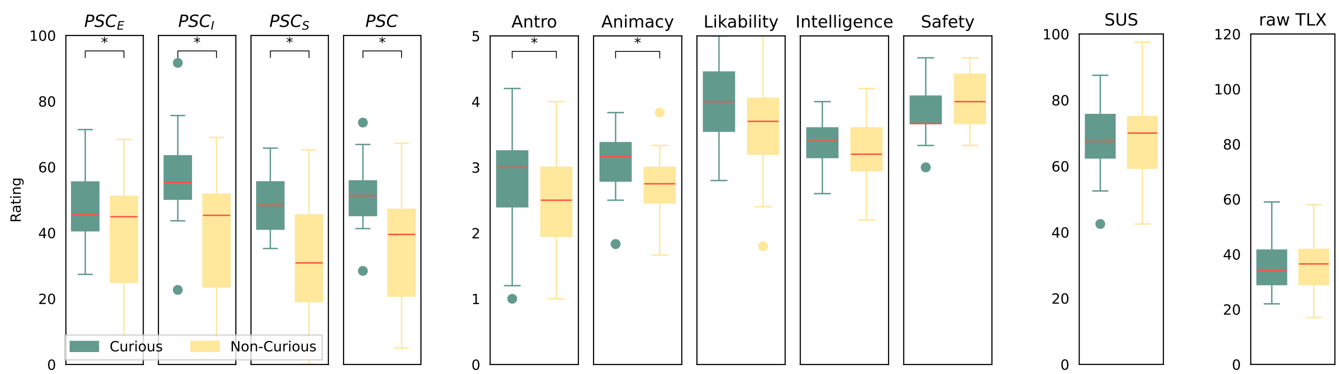
Figure 6: Boxplots comparing system and user behavior metrics between the *curious* (green) and *non-curious* (yellow) characters. The plots show distributions of positive sentiment utterances, actions per turn, and actor turns per minute, highlighting key differences in sentiment expression and behavior dynamics.

investigative, social) as the *non-curious* system and we found no statistically significant difference between the Scenarios.

5.2.2 Godspeed. The data for all godspeed subscales (*anthromopomorphism, animacy, likeability, perceived intelligence*) except for *perceived safety* are normally distributed. There was a significant main effect of Character on *anthropomorphism*. We did not find a significant main effect for Scenario and no interaction effects. There was a significant main effect of Character on *animacy*. There was not a significant main effect on Scenario and no interaction effects. There were no significant main effects of Character or Scenario on *likeability*, but we found significant interaction effects. We conducted t-tests on both scenarios independently to find whether the Character had an effect on *likeability* for each scenario and found a statistically significant effect for the *cocktail* Scenario

Table 2: Results of the statistical analysis of all questionnaires.

	Normality		CHARACTER			SCENARIO			S × C		
	W	**p**	**F**	**p**	η^2	**F**	**p**	η^2	**F**	**p**	η^2
SUS [9]	.983	.808	1.193	.112	.002	1.194	.289	.047	1.011	.328	.014
raw NASA-TLX [29]	.976	.526	.006	.938	<.001	.659	.427	.029	9.861	<.01*	.087
PSC											
Expressive	.971	.399	7.944	<.05*	.084	.148	.705	.006	.694	.416	.008
Investigative	.945	.051	14.896	<.001*	.242	.304	.588	.01	3.465	.079	.069
Social	.954	.104	28.666	<.001*	.296	.005	.946	<.001	.956	.341	.014
PSC	.951	.08	21.398	<.001*	.240	.002	.962	<.001	2.268	.149	.032
Goodspeed [4]											
Anthropomorphism	.972	.402	4.52	<.05*	.037	.132	.72	.006	.984	.334	.008
Animacy	.966	.274	10.166	<.01*	.116	.638	.435	.026	.189	.669	.002
Likeability	.97	.357	4.085	.058	.049	.011	.916	<.001	5.208	<.05*	.062
Perceived Intelligence	.974	.463	.474	.5	.013	.464	.504	.013	.303	.588	.008
Perceived Safety	.922	.009	4.15	.057	.187	.441	.515	.024	.416	.527	0.023



(a) PSC **(b) Godspeed** **(c) SUS** **(d) NASA-TLX**

Figure 7: Boxplots of questionnaire results per CHARACTER. From left to right: (a) PSC with the three subscales: PSC_E, PSC_I, PSC_S, (b) Godspeed with the five subscales: anthropomorphism, animacy, likability, perceived intelligence, perceived safety, (c) SUS, and (d) raw NASA-TLX.

($t(9) = 2.906, p < .05$), but not for *pizza* ($t(9) = -.194, p = .850$). We did not find any main or interaction effects of CHARACTER and SCENARIO for *perceived intelligence* or *perceived safety*. See Table 2 and Figure 7b. Thus, the *curious* system was perceived as significantly more *anthropomorph* and *animate*.

5.2.3 SUS. We found no statistically significant effects for SUS; see Table 2. The overall score was high with a mean of 69.125 for the *curious* condition and 68.250 for the *non-curious* condition; see Figure 7c. Indicating that both the *curious* and the *non-curious* systems had an overall "good" score based on Brooke [10].

5.2.4 Raw NASA-TLX. Our results of the raw NASA-TLX are normally distributed. We found that the main effects CHARACTER and SCENARIO are not statistically significant; see Table 2 and Figure 7d. However, we found statistically significant interaction effects. We conducted t-tests on both scenarios independently to find whether the CHARACTER had an effect on *perceived workload* for each scenario and found a statistically significant effect for the *cocktail* SCENARIO ($t(9) = -2.272, p < .05$), but not for *pizza* ($t(9) = 2.169, p = .582$).

Moreover, none of the individual raw NASA-TLX questions results are normally distributed data. Thus, we conducted ART ANOVAs to examine the effect of CHARACTER and SCENARIO on *mental demand*, *physical demand*, *temporal demand*, *performance*, *effort*, and *frustration*. We did not find any significant main or interaction effects for all questions except for *effort*. For *effort*, there was a significant main effect of CHARACTER ($F(1, 18) = 5.138$, $p < .05$), indicating that participants perceived more effort interacting with the *curious* system ($M = 6.40$) than the *non-curious* system ($M = 4.85$), no significant main effect of SCENARIO ($F(1, 18) = 1.138$, $p = .300$), and a significant interaction effect between CHARACTER and SCENARIO ($F(1, 18) = 4.586$, $p < .05$).

5.3 Interview Results

In the following, we present the results of the interviews according to the identified themes.

Preferred System. We asked all participants which system they preferred. Most (70%) preferred interacting with the curious system, while 20% preferred the non-curious system, and 10% had no preference. Participants who preferred the curious system noted that it felt more empathetic (P18) and frequently mentioned that they enjoyed that the system felt more interactive and natural:

> "*It was more personal. It interacted more, in a way, more humanistic [...] Even though it's a robot, it felt more natural*" (P9)

Those who preferred the non-curious system mentioned that the additional questions asked by the curious system were inefficient and distracting from the main goal of the task, while the non-curious system, "*in terms of achieving the goal, is more efficient*" (P4). One participant noted that, in general, he interacts with robots to accomplish tasks rather than to be social:

> "*I'm not interacting with the robot to be friendly with them. I'm interacting with the robot to get something done.*" (P1)

Perceived Difference Between Systems. We asked participants to identify the main differences they perceived between the two systems at the beginning of the interview. We identified nine themes, which we grouped into three categories: *human-like*, *curious*, and *autonomous*, shown in Table 3. The themes are framed positively from the perspective of the curious system (i.e., *the main difference is that the curious system is more X*). The *human-like* category incorporates social and anthropomorphic elements, the *curious* category includes specifically curious or inquisitive behaviors, and the *autonomous* includes references to the system acting on its own and explaining its actions.

Table 3: The main differences between the systems reported by the participants at the beginning of the interview. The themes are framed from the perspective of the curious system (i.e., *The the main difference is that the curious system is more X*).

Category	Theme	Count	Total
Human-Like	Social	6	16
	Human	4	
	Interactive	5	
	Friendly	1	
Curious	Questioning	5	9
	Explorative	3	
	Curious	1	
Autonomous	Explanatory	5	9
	Proactive	4	

Perceived Curiosity. Nearly all participants (90%) perceived that the curious system was more curious than the non-curious one. One participant perceived the non-curious system as more curious, and one could not differentiate. In the case of the one participant who perceived the non-curious system as more curious, their non-curious condition had an uncharacteristically high number of scenarios where the robot could not reach an object. These scenarios led to the non-curious system asking the participant for help on multiple occasions.

Observed Curious Behaviors. The behaviors that participants identified as curious can be broadly categorized into three groups. Social behaviors were mentioned by 90% of participants, while 60% noted perceptual exploration and 40% mentioned task-related questions.

Social interactions were the most frequently mentioned curious behavior. The participants noted that the system asked non-task-related social questions, saying that it "*wants to know my name, it wants to know my preferences*" (P2). These social interactions were interpreted as "*much more natural in conversation*" (P9).

Most participants also noted that the system appeared curious when it explored the environment perceptually. In particular, participants noted moments when the robot was visually inspecting the environment, saying that the "*gaze behavior was more active [...] it gave more of an impression that it's scanning the environment, considering*" (P13). They also mentioned instances when the robot poked or shook objects to learn more about their contents:

> "*He immediately was interested [in] what's in the hidden container to see what's inside there, and really pointing at it, shaking it.*" (P10)

Finally, participants also noted that the system appeared to be curious when it asked task-related questions, such as inquiring about objects or identifying unknowns in the environment. For example, participants noted that the system appeared curious when it "*asked what was in the container*" (P5) Some participants perceived this questioning to be tied to a desire to learn, saying "*it tries to learn, tries to adapt, and it also asks questions*" (P4).

In all, social interactions, perceptual exploration, and asking task-related questions were the most frequently noticed curious behaviors.

Properties of a Curious System. Beyond individual behaviors, participants perceived characteristic properties that were indicative of a curious system. Specifically, participants most frequently mentioned that the system appeared curious when they perceived that it was motivated to learn by "*exploring the environment*" (P4) and unknowns:

> "*It went for the unknown object by itself without me having to ask [...] It also interacted with the novel object.*" (P10)

Participants also perceived the properties of asking questions and trying to learn from the user as curious. They reported that the system "*was trying to learn from me*" (P9). Overall, they noted that when the system asked them questions, it felt more human:

> "*He was asking me to do something or to explain it to him, and it was more interactive this way, so it felt like you have kind of a second human.*" (P7)

Finally, the system was perceived as curious when it behaved proactively. The participants noticed that the curious system "*showed more initiative*" (P13) and "*was a lot more proactive*" (P1).

Increasing Perceived Curiousity. We asked participants how they imagined the system could be changed to increase the perception of curiosity. The responses generally related to increasing the behaviors and properties in the previous sections. For example, the participants suggested asking more questions (P15), interacting with more objects (P2), exploring more of the environment (P17), and being more proactive (P5). On a higher level, participants suggested that a system would appear more curious if it learned and improved over time, displaying a "*willingness to learn combined with success in learning something*" (P3). Finally, some participants noted that it may have been easier for the system to express curiosity in a more complex or uncertain task.

Interaction Modality. We also asked participants to reflect on the different interaction modalities present in the system. Most participants noted that both verbal and non-verbal modalities were useful for explainability and helping the user to understand the intentions of the system:

> "*Verbal communication is the most important one in this case because if it doesn't say anything, I don't know what's happening.*" (P5)
> "*He looked at where he wanted to grasp. That was, I think, good in terms of me knowing what he's doing.*" (P19)

Participants also noticed that different modalities can be advantageous in different contexts and tasks. Verbal interaction can be useful to quickly learn information from users, saying "*It could just ask me if I know what is in the box, which would be also curious behavior; it's more efficient*" (P19). Conversely, non-verbal behaviors are appropriate when the system needs to try things out and explore to learn, noting "*in the wild there are tasks where you actually have to try out things [...] to actually interact with objects*" (P11). On the other hand, they noted that physical exploration in unknown environments could lead to dangerous situations:

> "*Would you want the robot to shake the pan? Maybe dripping hot oil or something, right?*" (P13)

Long Term Use. Finally, we asked participants to think forward and imagine benefits and concerns about curious robots in the future, and to imagine use cases where they would want to interact with a curious robot. The most frequently mentioned benefits related to learning. Participants noted that "*the advantage of curiosity is that over time, the robot should learn more*" (P11). They noted that this would result in more questions at the beginning but would make it "*make it efficient in the long term*" (P12). Participants also imagined that the system and the users could learn together:

> "*If there's a task that I don't know yet, and also the robot doesn't know [...] we both can learn something about a task.*" (P11)

Conversely, participants also expressed some concerns about a curious system. They imagined that the system could be annoying (P12), noting that it would be important for the system to understand "*when curiosity is okay and when it is not okay*" (P13). Ultimately, participants gave the impression that it is more important for the system to be effective, whether it learns tasks by being "*curious, or whether it's just machine learned in advance*" (P2). Although some participants noted privacy concerns (P13), others suggested that systems could share information to grow their intelligence collectively:

> "*Share their experiences across each other, across systems distributed around laboratories or people's homes*" (P2)

Regarding use cases, participants most frequently mentioned that they would want a curious system for learning unfamiliar tasks where both the system and the user are gaining knowledge:

> "*In a situation where it's helping me discover, like acquire knowledge. So, in an exploration setting, when we are exploring an environment together.*" (P1)

Many participants also noted that a curious system would be useful for household tasks, as it would learn their preferences over time. Lastly, participants frequently mentioned social tasks, such as interacting with elderly (P16) or lonely (P17) people, where a curious system would be beneficial.

6 Discussion

Based on the results from our user study, in which 20 participants interacted with a curious and a non-curious autonomous system, we discuss our findings on creating a perceivable curious robot, how system curiosity affects user experience, and discuss using MM-LLMs to modulate system behavior. Lastly, we provide recommendations on how to design system curiosity.

6.1 Users Can Perceive the Robots' Curiosity in a Collaborative Task

Both through the qualitative data and questionnaire results, we found that participants perceived the *curious* CHARACTER as more curious, indicating that our manipulation was successful. First, in direct comparison and without any priming, 90% of participants stated when asked to compare that they found the *curious* system more curious than the *non-curious* system. While we did not explicitly create curious behaviors for different types of curiosity, we found that participants named aspects of the system that we can now categorize. Furthermore, we found that participants always perceived the *curious* system as significantly more socially curious, investigative, and explorative through the three subscales of the PSC. This is also supported by the interviews, in which the majority of participants stated that they noticed the socially curious and explorative behavior most often. The investigativeness of the system was also named in the interviews, as participants noticed both task-related and task-unrelated questions as curious questions. Thus, we could successfully create user-centric curiosity for an MM-LLM-driven robot.

Previous works primarily created curious robot behaviors for robot learning [13, 51], educational tasks [14, 58], or off-task behavior [70]. Conversely, we created an autonomous system, which decided when and how to be curious during collaborative service tasks. Our findings that curious behavior is observable align with the findings from previous work [22, 70]. Importantly, we tested this system across two different settings and did not find any differences in perceived curiosity, showing the generalizability of our system across tasks. We found that the two CHARACTERS interacted with the user in a significantly different way. To summarize, we showed that we can change the perceived curiosity in a service robot by giving the autonomous system a character via an LLM system prompt.

6.2 Curious System Behavior Improves User Experience

When designing the system prompts to modulate the robot's character and behavior, we aimed to vary its perceived curiosity without altering its intelligence or other capabilities. Although we observed no significant differences in perceived intelligence between conditions, this does not definitively isolate curiosity as the sole factor influencing user perceptions. Nevertheless, our findings indicate that introducing curiosity into the system's behavior led to significant positive effects for anthropomorphism and animacy. Furthermore, the most stated difference between the two systems by the participants was that the *curious* system was more human-like, social, and interactive. Thus, we found that curious system behavior also increases human-likeness and lifelikeness/animacy. This is supported by literature giving other human-like traits to technical systems [54]. Previous work demonstrates that systems exhibiting human-like traits foster greater willingness to interact, leading to enhanced human-robot interaction [3, 39, 54, 74]. Robot curiosity has also been shown to help spark human curiosity and, thus, increases human interest in the task [14, 58]. Thus, giving a system curiosity as a trait that changes its behavior can make this interaction more engaging for users.

Walker et al. [70] found that off-task curiosity might be negatively perceived but also stated that this can be mitigated by letting the robot acknowledge and explain its decision process. In our collaborative setting, the robot would usually explain verbally what it planned to do, hence proactively giving an insight into its reasoning. Many participants also stated the importance of understanding the actions and decisions of the robot, either from the robot verbally supporting its actions or through transparent non-verbal actions.

6.3 Using MM-LLMs to Create and Modulate Personalities for Systems

Previously, with model- and rule-based systems, it was difficult to rapidly create certain behaviors for robots. Model-based systems can trigger single curious behaviors driven by a reward function [70]. However, such behaviors require delicately designed reward functions. Furthermore, in such systems, the model only triggers the curious behavior; the dialog and interaction flow still might be missing or rely on rule-based templates, which leads to scenario-dependent interaction designs. Previous research proved that LLMs can successfully generate expressions [47] and adjust their actions according to human input [2, 78]. Other have also used LLMs to create expressive robot actions when interacting with users [65, 72]. However, previous work still has not shown whether LLMs can be used to regulate robot behavior according to a certain character-level trait *across different tasks*. In our study, we show that we were able to create one system prompt, which was able to drive interaction across two different tasks, indicating the possibility for more generalizability. However, currently, it is still necessary to engineer the prompt in a way that works for all used tasks.

When setting the curious system character, we aimed for the robot to exhibit a wide range of behaviors. Every interaction with the robot triggered some of these curious behaviors. However, which and how often each behavior occurred depended on how the user interacted with the system. In some interactions, the user would already state information the robot needed, leading to the robot not performing the expressive curious behavior we implemented for that situation but rather relying on the verbally gathered knowledge. This shows that we could use the LLM to imbue the system with a persona instead of just providing it with a set of curious capabilities. This is in line with Schmidt et al. [62], who outline how LLMs can be used to create living characters from personas that users can interact with in real time.

In our manipulation checks, we found that both systems used all capabilities, and especially that the curious system would use the capabilities in a way to express user-centric curiosity, showing that we can change the system's persona, which is in line with related work [30]. We originally aimed to provide this persona only on a high level and hoped that the system would, via this trait, perform the expected behaviors correctly itself. However, currently, LLMs still suffer from forgetfulness [16], which led us to structure the system prompt into three sections, in which we had to fine-tune the prompt by giving explicit instructions, similar to if-statements. Furthermore, another improvement to the general system interaction flow was using compound actions instead of atomic actions, which led to the LLM needing to call less independent actions. Future work should look into giving even less direct guidance via the system prompt, which would generalize the curious persona even more to any real-world setting.

6.4 Limitations

Participants interacted with our curious system for a single task, so we did not investigate long-term effects. However, we believe that long-term interaction is where curiosity might have its strongest advantages. Yet, we found that even in this short task duration, participants not only noticed the curiosity, but the majority even preferred interacting with the curious system due to its higher human likeness and interactivity. Future work should investigate the effect of curiosity in self-learning systems on user perception.

Furthermore, many participants stated that one factor limiting the system from appearing even more curious was the relative simplicity of the task. As our study is a first attempt to understand how curious systems affect user experience; we created a controlled task with limited items. This ensured a fair level of reproducibility across participants. In the future, more complex or uncertain tasks may provide more opportunities for the robot to express curiosity.

Robots are currently still novel to most people, and we wanted to study the effect of different robot characters on user perception. As we did not want the study to be biased by the novelty effect of people seeing a moving robot for the first time and rather focused on the interaction with the robot, we decided to recruit the majority of our participants with experience with robots. However, this is still a limited sample of the general population, even when robots become more ubiquitous, and thus, future work should look at this further.

Lastly, we used an autonomous system with minimal limitations for the user. Due to the nature of the open system, it could happen that depending on how a participant behaves, they may have a noticeably different experience. In our case, this led to one participant even perceiving the non-curious system as more curious because, in their case, the system behaved differently than we expected.

7 Design Recommendations for Curious Systems

From the PSC questionnaire results and interviews, we showed that we could increase the perceived curiosity of our system. From our system design and the interviews, we propose the following design recommendations, which future system designers should encompass when creating curious behavior for interactive systems.

Create socially curious behavior for higher user acceptance. Socially curious behaviors, such as asking the user for their name or preferences, were very noticeable and appreciated by users. While not essential for task completion, participants perceived the system as more friendly, human-like, and interactive because of the social behavior, leading to a more natural interaction. Through social curiosity, a system can learn more about the user and, thus, enhance personalization. Furthermore, social curiosity fosters closeness in human interaction [33] and, combined with anthropomorphism, increases trust in systems [20, 50], which can improve human-robot team performance [17].

Ask task-related questions to make the system appear curious. When humans do not know something, they typically ask for help from another person who might know the answer. Oftentimes, this can be the most efficient way to obtain the correct, contextually relevant information. In robot systems, where efficiency is an important factor, the same approach should be applied. In our study, participants interpreted task-related questions as the system having a desire to learn. In an open-world scenario, where efficiency plays an even greater role, task-related questions can serve as a practical and adaptive way for the system to gather critical information about its surroundings. Participants highlighted that increasing the frequency and depth of such questions would be a key improvement, as it would further enhance the system's perceived curiosity and engagement. This suggests that actively involving users through inquiry not only helps the system learn but also fosters a more interactive and dynamic user experience.

Use explorative non-verbal behavior to communicate perceptual curiosity. Our system used non-verbal methods, such as shaking (auditive), poking (haptic), and observing (visual), to explore the environment. Participants named these gestures most often when we asked them which system behaviors they perceived as curious, which is in line with our findings of the *curious* system being significantly more curious than the *non-curious* system. Such gestures reflect perceptual curiosity, essential for reducing uncertainty and gathering new information [67], especially in open-world scenarios where users may not have the answers, requiring the system to gather information independently. Participants also reported that the combination of verbal and non-verbal communication enhanced the explainability of the system.

Employ the inherent proactivity in curious systems to contribute to collaboration. Curiosity arises from detecting knowledge gaps [45], prompting one to take actions to reduce uncertainty. Similarly, supportive systems should proactively explore and address information gaps to gain knowledge. Our participants identified proactivity and autonomy as key differences between our *curious* and *non-curious* systems and appreciated the system investigating and proposing potential next steps for the task itself. Integrating this proactivity may not be appropriate in every system depending on the intended interaction flow, but ultimately it can positively contribute to user experience by making the robot an active partner.

Make the learning benefits of curiosity apparent in long-term interactions. While most participants preferred interacting with the curious system, some prioritized efficiency. Interaction with the curious system required significantly more effort, measured through the raw NASA-TLX. Furthermore, this initial interaction with the curious system led to a longer task duration due to the additional interactions. To reduce this additional effort for long-term use, curiosity also must lead to increased knowledge, reducing the amount of needed interactions with the user. This should lead to more efficient interactions over time, reducing the user's burden. This learning effect should be apparent to the user so that they understand the benefits associated with the additional initial efforts.

Balance user-centric curiosity with task-efficiency. When designing a curious system, it is necessary to balance when and how often the system expresses curious behaviors. We expect that robots will primarily be used as assistive systems in the future, meaning their main job is to assist with completing tasks. While a curious system will become more efficient over time, and curiosity is also needed as it is impossible to teach a system *everything* in development, it is still important to balance curious behaviors with task completion so that users are not overwhelmed, and efficiency is maintained. Moreover, users should be able to control and balance the level and type (social, epistemic, perceptual) of curiosity expressed by the robot depending on the context. Aligning these behaviors with users' perceptions of the robot's capabilities is critical, as a mismatch—whether through under-perception or over-perception of the system's actual abilities—can significantly impact user acceptance and trust in the system [15].

8 Conclusion

In this work, we successfully created measurable curious behavior for a robot in a collaborative task. We developed and used an open loop MM-LLM system, with which we could tune the robot's character to be either curious or non-curious. In a user study ($N = 20$), we found that users found the curious character significantly more curious, anthropomorphic, animate, autonomous, and interactive. Based on our findings, we propose design recommendations on how future work can design curious behavior. We also show that not only the user perception was affected, but we could objectively change the behavior of the system. We propose that with this system, future work can use expert knowledge of other domains and build different modulations of this personality trait for a robot.

References

[1] Ferran Alet, Martin F Schneider, Tomas Lozano-Perez, and Leslie Pack Kaelbling. 2020. Meta-learning curiosity algorithms. *arXiv preprint arXiv:2003.05325* (2020).

[2] Philipp Allgeuer, Hassan Ali, and Stefan Wermter. 2024. When robots get chatty: Grounding multimodal human-robot conversation and collaboration. *arXiv preprint arXiv:2407.00518* (2024).

[3] Alexander Mois Aroyo, Francesco Rea, and Alessandra Sciutti. 2017. Will You Rely on a Robot to Find a Treasure?. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna, Austria) *(HRI '17)*. Association for Computing Machinery, New York, NY, USA, 71–72. doi:10.1145/3029798.3038394

[4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1 (2009), 71–81.

[5] Anna Belardinelli, Chao Wang, Daniel Tanneberg, Stephan Hasler, and Michael Gienger. 2024. Train your robot in AR: investigating user experience with a continuously learning robot. (2024). https://leap-hri.github.io/docs/camready_4.pdf

[6] Erdem Biyik, Malayandi Palan, Nicholas C. Landolfi, Dylan P. Losey, and Dorsa Sadigh. 2019. Asking Easy Questions: A User-Friendly Approach to Active Reward Learning. *ArXiv* abs/1910.04365 (2019). https://api.semanticscholar.org/CorpusID:204008187

[7] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI Research: Going Behind the Scenes. *Synthesis Lectures on Human-Centered Informatics* 9, 1 (April 2016), 1–115. doi:10.2200/S00706ED1V01Y201602HCI034

[8] Cynthia Breazeal and Brian Scassellati. 1999. How to build robots that make friends and influence people. In *Proceedings 1999 IEEE/RSJ international conference on intelligent robots and systems. Human and environment friendly robots with high intelligence and emotional quotients (cat. No. 99CH36289)*, Vol. 2. IEEE, 858–863.

[9] John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[10] John Brooke. 2013. SUS: a retrospective. *Journal of Usability Studies* 8 (01 2013), 29–40.

[11] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A. Efros. 2018. Large-Scale Study of Curiosity-Driven Learning. arXiv:1808.04355 [cs.LG] https://arxiv.org/abs/1808.04355

[12] Maya Cakmak, Crystal Chao, and Andrea L. Thomaz. 2010. Designing Interactions for Robot Active Learners. *IEEE Transactions on Autonomous Mental Development* 2, 2 (2010), 108–118. doi:10.1109/TAMD.2010.2051030

[13] Maya Cakmak and Andrea L Thomaz. 2012. Designing robot learners that ask good questions. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. IEEE, 17–24.

[14] Jessy Ceha, Nalin Chhibber, Joslin Goh, Corina McDonald, Pierre-Yves Oudeyer, Dana Kulić, and Edith Law. 2019. Expression of Curiosity in Social Robots: Design, Perception, and Effects on Behaviour. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300636

[15] Elizabeth Cha, Anca D. Dragan, and Siddhartha S. Srinivasa. 2015. Perceived robot capability. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 541–548. doi:10.1109/ROMAN.2015.7333656

[16] Juo-Tung Chen and Chien-Ming Huang. 2023. Forgetful large language models: Lessons learned from using LLMS in robot programming. In *Proceedings of the AAAI Symposium Series*, Vol. 2. 508–513.

[17] Min Chen, Stefanos Nikolaidis, Harold Soh, David Hsu, and Siddhartha Srinivasa. 2018. Planning with Trust for Human-Robot Collaboration. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Chicago IL USA, 307–315. doi:10.1145/3171221.3171264

[18] Rohan Chitnis, Tom Silver, Joshua B Tenenbaum, Leslie Pack Kaelbling, and Tomás Lozano-Pérez. 2021. Glib: Efficient exploration for relational model-based reinforcement learning via goal-literal babbling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 11782–11791.

[19] Robert P Collins, Jordan A Litman, and Charles D Spielberger. 2004. The measurement of perceptual curiosity. *Personality and Individual Differences* 36, 5 (2004), 1127–1141. doi:10.1016/S0191-8869(03)00205-8

[20] Ewart J. De Visser, Samuel S. Monfort, Ryan McKendrick, Melissa A. B. Smith, Patrick E. McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents. *Journal of Experimental Psychology: Applied* 22, 3 (Sept. 2016), 331–349. doi:10.1037/xap0000092

[21] Nikos Dimitropoulos, Pantelis Papalexis, George Michalos, and Sotiris Makris. 2024. Advancing Human-Robot Interaction Using AI – A Large Language Model (LLM) Approach. In *Advances in Artificial Intelligence in Manufacturing*, Achim Wagner, Kosmas Alexopoulos, and Sotiris Makris (Eds.). Springer Nature Switzerland, Cham, 116–125.

[22] Malcolm Doering, Phoebe Liu, Dylan F. Glas, Takayuki Kanda, Dana Kulić, and Hiroshi Ishiguro. 2019. Curiosity Did Not Kill the Robot: A Curiosity-based Learning System for a Shopkeeper Robot. *J. Hum.-Robot Interact.* 8, 3, Article 15 (jul 2019), 24 pages. doi:10.1145/3326462

[23] Danny Driess, Fei Xia, Mehdi S. M. Sajjadi, Corey Lynch, Aakanksha Chowdhery, Brian Ichter, Ayzaan Wahid, Jonathan Tompson, Quan Vuong, Tianhe Yu, Wenlong Huang, Yevgen Chebotar, Pierre Sermanet, Daniel Duckworth, Sergey Levine, Vincent Vanhoucke, Karol Hausman, Marc Toussaint, Klaus Greff, Andy Zeng, Igor Mordatch, and Pete Florence. 2023. PaLM-E: An Embodied Multimodal Language Model. (3 2023). http://arxiv.org/abs/2303.03378

[24] Lisa A. Elkin, Matthew Kay, James J. Higgins, and Jacob O. Wobbrock. 2021. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. 754–768 pages. doi:10.1145/3472749.3474784

[25] Julia Fink. 2012. Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In *Social Robotics*, Shuzhi Sam Ge, Oussama Khatib, John-John Cabibihan, Reid Simmons, and Mary-Anne Williams (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–208.

[26] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292. doi:10.1016/j.patcog.2014.01.005

[27] Goren Gordon. 2019. Social behaviour as an emergent property of embodied curiosity: a robotics perspective. *Philosophical Transactions of the Royal Society B* 374, 1771 (2019), 20180029.

[28] Goren Gordon, Cynthia Breazeal, and Susan Engel. 2015. Can Children Catch Curiosity from a Social Robot?. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, Oregon, USA) *(HRI '15)*. Association for Computing Machinery, New York, NY, USA, 91–98. doi:10.1145/2696454.2696469

[29] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. doi:10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[30] 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* (Mexico City, Mexico, 2024). Association for Computational Linguistics, 3605–3627. doi:10.18653/v1/2024.findings-naacl.229

[31] Frank Joublin, Antonello Ceravola, Pavel Smirnov, Felix Ocker, Joerg Deigmoeller, Anna Belardinelli, Chao Wang, Stephan Hasler, Daniel Tanneberg, and Michael Gienger. 2024. CoPAL: corrective planning of robot actions with large language models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 8664–8670.

[32] Min Jeong Kang, Ming Hsu, Ian M Krajbich, George Loewenstein, Samuel M McClure, Joseph Tao-yi Wang, and Colin F Camerer. 2009. The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological science* 20, 8 (2009), 963–973.

[33] Todd B. Kashdan, Patrick E. McKnight, Frank D. Fincham, and Paul Rose. 2011. When Curiosity Breeds Intimacy: Taking Advantage of Intimacy Opportunities and Transforming Boring Conversations. *Journal of Personality* 79, 6 (Dec. 2011), 1369–1402. doi:10.1111/j.1467-6494.2010.00697.x

[34] Todd B Kashdan and John E Roberts. 2004. Trait and state curiosity in the genesis of intimacy: Differentiation from related constructs. *Journal of Social and Clinical Psychology* 23, 6 (2004), 792–816.

[35] Todd B Kashdan, Ryne A Sherman, Jessica Yarbro, and David C Funder. 2013. How are curious people viewed and how do they behave in social situations? From the perspectives of self, friends, parents, and unacquainted observers. *Journal of personality* 81, 2 (2013), 142–154.

[36] Matthew Kay, Lisa A. Elkin, James J. Higgins, and Jacob O. Wobbrock. 2021. *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs*. doi:10.5281/zenodo.594511 R package version 0.11.1.

[37] Celeste Kidd and Benjamin Y Hayden. 2015. The psychology and neuroscience of curiosity. *Neuron* 88, 3 (2015), 449–460.

[38] 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* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 371–380. doi:10.1145/3610977.3634966

[39] Dieta Kuchenbrandt, Friederike Eyssel, Simon Bobinger, and Maria Neufeld. 2011. Minimal Group - Maximal Effect? Evaluation and Anthropomorphization of the Humanoid Robot NAO. In *Social Robotics*, Bilge Mutlu, Christoph Bartneck, Jaap Ham, Vanessa Evers, and Takayuki Kanda (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 104–113.

[40] Edith Law, Vicky Cai, Qi Feng Liu, Sajin Sasy, Joslin Goh, Alex Blidaru, and Dana Kulic. 2017. A Wizard-of-Oz Study of Curiosity in Human-Robot Interaction. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, Lisbon, 607–614. doi:10.1109/ROMAN.2017.8172365

[41] Michael A. Lawrence. 2016. Ez: Easy Analysis and Visualization of Factorial Experiments.

[42] Jan Leusmann, Steeven Villa, Burak Berberoglu, Chao Wang, and Sven Mayer. 2025. Developing and Validating the Perceived System Curiosity Scale (PSC): Measuring Users' Perceived Curiosity of Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI ́25)*. Association for Computing Machinery, New York, NY, USA. doi:10.1145/3706598.3713087

[43] Filip Lievens, Spencer H Harrison, Patrick Mussel, and Jordan A Litman. 2022. Killing the cat? A review of curiosity at work. *Academy of Management Annals* 16, 1 (2022), 179–216.

[44] Jordan A. Litman. 2012. *Epistemic Curiosity*. Springer US, Boston, MA, 1162–1165. doi:10.1007/978-1-4419-1428-6_1645

[45] George Loewenstein. 1994. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin* 116, 1 (1994), 75.

[46] Kurosh Madani, Christophe Sabourin, and Dominik M Ramík. 2013. Artificial curiosity emerging human-like behavior: Toward fully autonomous cognitive robots. In *International Joint Conference on Computational Intelligence*. Springer, 501–516.

[47] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. 2024. Generative Expressive Robot Behaviors using Large Language Models. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 482–491. doi:10.1145/3610977.3634999

[48] Cynthia Matuszek, Nick Depalma, Ross Mead, Tom Williams, and Ruchen Wen. 2024. Scarecrows in Oz: Large Language Models in HRI. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 1338–1340. doi:10.1145/3610978.3638168

[49] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. 2020. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics* 12, 6 (2020), 1179–1201.

[50] Manisha Natarajan and Matthew Gombolay. 2020. Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Cambridge United Kingdom, 33–42. doi:10.1145/3319502.3374839

[51] Sao Mai Nguyen, Serena Ivaldi, Natalia Lyubova, Alain Droniou, Damien Gérardeaux-Viret, David Filliat, Vincent Padois, Olivier Sigaud, and Pierre-Yves Oudeyer. 2013. Learning to recognize objects through curiosity-driven manipulation with the iCub humanoid robot. In *2013 IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. 1–8.

[52] Olivia Nocentini, Laura Fiorini, Giorgia Acerbi, Alessandra Sorrentino, Gianmaria Mancioppi, and Filippo Cavallo. 2019. A survey of behavioral models for social robots. *Robotics* 8, 3 (2019), 54.

[53] Felix Ocker, Jörg Deigmöller, and Julian Eggert. 2023. Exploring Large Language Models as a Source of Common-Sense Knowledge for Robots. *arXiv preprint arXiv:2311.08412* (2023).

[54] Benjamin Oistad, Catherine Sembroski, Kathryn Gates, Margaret Krupp, Marlena Fraune, and S. Sabanovic. 2016. Colleague or Tool? Interactivity Increases Positive Perceptions of and Willingness to Interact with a Robotic Co-worker, Vol. 9979. 774–785. doi:10.1007/978-3-319-47437-3_76

[55] Pierre-Yves Oudeyer. 2018. Computational theories of curiosity-driven learning. *arXiv preprint arXiv:1802.10546* (2018).

[56] Deepak Pathak, Pulkit Agrawal, Alexei A Efros, and Trevor Darrell. 2017. Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning*. PMLR, 2778–2787.

[57] Britta Renner. 2006. Curiosity About People: The Development of a Social Curiosity Measure in Adults. *Journal of Personality Assessment* 87, 3 (2006), 305–316. doi:10.1207/s15327752jpa8703_11

[58] Maor Rosenberg, Hae Won Park, Rinat Rosenberg-Kima, Safinah Ali, Anastasia K. Ostrowski, Cynthia Breazeal, and Goren Gordon. 2021. Expressive Cognitive Architecture for a Curious Social Robot. *ACM Trans. Interact. Intell. Syst.* 11, 2, Article 12 (jul 2021), 25 pages. doi:10.1145/3451531

[59] Debra L. Roter, Susan M. Larson, Mary Catherine Beach, and Lisa A. Cooper. 2008. Interactive and evaluative correlates of dialogue sequence: A simulation study applying the RIAS to turn taking structures. *Patient Education and Counseling* 71, 1 (2008), 26–33. doi:10.1016/j.pec.2007.10.019

[60] Maha Salem, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2011. A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction. In *IEEE International Workshop on Robot and Human Communication (RO-MAN)*. IEEE, 247–252. doi:10.1109/ROMAN.2011.6005285

[61] Jürgen Schmidhuber. 1991. Curious model-building control systems. In *Proc. international joint conference on neural networks*. 1458–1463.

[62] Albrecht Schmidt, Passant Elagroudy, Fiona Draxler, Frauke Kreuter, and Robin Welsch. 2024. Simulating the Human in HCD with ChatGPT: Redesigning Interaction Design with AI. *Interactions* 31, 1 (jan 2024), 24–31. doi:10.1145/3637436

[63] Yunfan Shao, Linyang Li, Junqi Dai, and Xipeng Qiu. 2023. Character-LLM: A Trainable Agent for Role-Playing. arXiv:2310.10158 [cs.CL] https://arxiv.org/abs/2310.10158

[64] D.S. Syrdal, K. Dautenhahn, K.L. Koay, and M.L. Walters. 2009. The Negative Attitudes Towards Robots Scale and reactions to robot behaviour in a live Human-Robot Interaction study. In *Adaptive and Emergent Behaviour and Complex Systems*. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB), United Kingdom, 109–115.

[65] Daniel Tanneberg, Felix Ocker, Stephan Hasler, Joerg Deigmoeller, Anna Belardinelli, Chao Wang, Heiko Wersing, Bernhard Sendhoff, and Michael Gienger. 2024. To Help or Not to Help: LLM-based Attentive Support for Human-Robot Group Interactions. *arXiv preprint arXiv:2403.12533* (2024).

[66] Christiane Attig Thomas Franke and Daniel Wessel. 2019. A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction* 35, 6 (2019), 456–467. doi:10.1080/10447318.2018.1456150

[67] Lieke Lf Van Lieshout, Floris P De Lange, and Roshan Cools. 2020. Why so Curious? Quantifying Mechanisms of Information Seeking. *Current Opinion in Behavioral Sciences* 35 (Oct. 2020), 112–117. doi:10.1016/j.cobeha.2020.08.005

[68] Sai Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. 2023. Chatgpt for robotics: Design principles and model abilities. *Microsoft Auton. Syst. Robot. Res* 2 (2023), 20.

[69] Naoki Wake, Atsushi Kanehira, Kazuhiro Sasabuchi, Jun Takamatsu, and Katsushi Ikeuchi. 2023. Chatgpt empowered long-step robot control in various environments: A case application. *arXiv preprint arXiv:2304.03893* (2023).

[70] Nick Walker, Kevin Weatherwax, Julian Allchin, Leila Takayama, and Maya Cakmak. 2020. Human Perceptions of a Curious Robot that Performs Off-Task Actions. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, United Kingdom) *(HRI '20)*. Association for Computing Machinery, New York, NY, USA, 529–538. doi:10.1145/3319502.3374821

[71] 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 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 218, 10 pages. doi:10.1145/3613905.3651029

[72] 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 *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction* (Boulder, CO, USA) *(HRI '24)*. Association for Computing Machinery, New York, NY, USA, 1105–1109. doi:10.1145/3610978.3640562

[73] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '11)*. ACM Press, New York, 143–146. doi:10.1145/1978942.1978963

[74] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. 2005. "Lifelike" behavior of communication robots based on developmental psychology findings. In *5th IEEE-RAS International Conference on Humanoid Robots, 2005*. IEEE, 406–411. doi:10.1109/ICHR.2005.1573601

[75] Yang Ye, Hengxu You, and Jing Du. 2023. Improved trust in human-robot collaboration with ChatGPT. *IEEE Access* (2023).

[76] Takuma Yoneda, Jiading Fang, Peng Li, Huanyu Zhang, Tianchong Jiang, Shengjie Lin, Ben Picker, David Yunis, Hongyuan Mei, and Matthew R Walter. 2023. Statler: State-maintaining language models for embodied reasoning. *arXiv preprint arXiv:2306.17840* (2023).

[77] Takahide Yoshida, Atsushi Masumori, and Takashi Ikegami. 2023. From Text to Motion: Grounding GPT-4 in a Humanoid Robot "Alter3". (12 2023). http://arxiv.org/abs/2312.06571

[78] Xufeng Zhao, Mengdi Li, Cornelius Weber, Muhammad Burhan Hafez, and Stefan Wermter. 2023. Chat with the Environment: Interactive Multimodal Perception Using Large Language Models. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 3590–3596. doi:10.1109/IROS55552.2023.10342363