A Robot-Assisted Approach to Small Talk Training for Adults with ASD

Rebecca Ramnauth[1], Dražen Brščić[2], Brian Scassellati[1]
[1]Department of Computer Science, Yale University, USA
[2]Graduate School of Informatics, Kyoto University, Japan
rebecca.ramnauth@yale.edu

Abstract—From dating to job interviews, making new friends or simply chatting with the cashier at checkout, engaging in small talk is a vital, everyday social skill. For adults with Autism Spectrum Disorder (ASD), small talk can be particularly challenging, yet it is essential for social integration, building relationships, and accessing professional opportunities. In this study, we present our development and evaluation of an in-home autonomous robot system that allows users to practice small talk. Results from the week-long study show that adults with ASD enjoyed the training, made notable progress in initiating conversations and improving eye contact, and viewed the system as a valuable tool for enhancing their conversational skills.

Index Terms—human-robot interaction, autism spectrum disorder, socially assistive robotics, social-skills training

I. INTRODUCTION

Imagine a scene where three coworkers are engaging in small talk at the beginning of their workday. One of them is Alex, who has Autism Spectrum Disorder (ASD), a neurodevelopmental condition that often makes it challenging to understand and interpret social cues [2].

C: Hey everyone, how's it going?
B: Hi! Not bad... just trying to power through this Monday. How about you, Alex?
A: Monday is okay.
C: Good to hear. Anything exciting happen this weekend?
B: Yeah, I finally tried that new restaurant. It was fantas—
A: I watched a movie.
C: Glad you liked the restaurant, Ben. It's my kids' favorite spot these days... What movie did you watch, Alex?
A: "The Martian."
B: I love that one! Matt Damon is awesome.
A: *No response.*

At first glance, this brief example of a typical interaction appears unremarkable. It represents the everyday small talk that occurs regularly in many workplaces. For workers with ASD, however, such apparently "easy" interactions may present a real challenge. In this example, while Alex responds to direct questions, the responses are brief and lack elaboration. Alex's responses provide minimal information rather than actively participating in the flow of the conversation. Additionally, Alex's lack of response to the last prompt may suggest difficulty in extending or sustaining the dialogue.

Although workers with ASD are often highly trained and skilled in job-specific tasks, they frequently face challenges with social interactions in the workplace. Müller et al. [42]

and Hurlbutt & Chalmers [26] found that many of their participants with ASD, despite completing graduate-level coursework, were employed in positions for which they were overqualified, such as food services or low-level administrative roles. This underemployment is further highlighted in the National Longitudinal Transition Study [64].

However, the ability to perform work tasks is only the tip of the iceberg when it comes to workplace success. Interpersonal skills are proven to be more significant predictors of overall success [25, 29, 48]. Hurlbutt & Chalmers [26] found that workers with ASD often attributed job-related challenges to social factors rather than the work-specific tasks themselves. In their study, one interviewee noted, "Jobs usually are 80% social (conversation, lunch, breaks, chit-chat) and 20% work." Furthermore, adults with ASD have reported that difficulty engaging in "social niceties," such as small talk, led to feelings of isolation and alienation in the workplace [26, 42].

Unlike the functional aspects of other on-the-job communication, which may focus on conveying information or assistance, small talk is considered purely social. It acts as a social lubricant, fostering rapport, mutual understanding, and trust [12], and is widely recognized as a key facilitator in building and maintaining relationships. In professional settings, small talk is considered an essential tool for networking success and establishing positive first impressions [56, 8]. It is even regarded as a vital skill that should be targeted in communication therapy for various populations [25, 29, 20].

To better support adults with ASD, it is important to develop accessible and targeted opportunities for improving interpersonal skills, such as small talk. Social robotics has the potential to enhance existing training initiatives by improving access to personalized, socially-situated, and physically co-present interactions [39, 66, 10, 58]. Physically present robots have proven effective in improving users' social abilities [67, 58], providing benefits such as enhanced engagement, improved social confidence, and greater motivation to participate—outcomes that are notably more pronounced than those achieved with non-embodied technologies [43].

Furthermore, research has established that robots for ASD interventions can result in positive and productive outcomes [66]. Social robots have demonstrated general effectiveness in enhancing verbal communication skills [33, 62, 31], including the ability to engage in everyday dialogue [3, 27, 45]. Additionally, participants with ASD in recent studies have

described such robot-assisted training as relevant and useful in their workplace experiences [41, 58].

Thus, leveraging our prior successes in developing robots for ASD interventions [66, 67, 58], we developed an autonomous training system that helps adults with ASD improve their small talk skills. Given the literature on the communicative difficulties of ASD (Sec. II), we begin by examining the extent to which small talk is considered a desired social skill (Sec. III). These insights inform our design requirements (Sec. V) and guide the development of our prototype (Sec. IV). In that formative study, we investigate initial impressions, perceptions of robot-assisted training, and anticipated outcomes for users with ASD. We then present findings from a week-long, in-home deployment (Sec. VII), highlighting how users with ASD received and engaged with the robot-assisted training.

II. RELATED WORK

Atypical communicative behaviors are key diagnostic criteria for ASD [2], often presenting as limited eye contact, difficulty understanding sarcasm or abstract language [55], and challenges in grasping the social rules that govern everyday interactions [70]. However, everyday, casual conversations are a pervasive aspect of daily life, whether it is chatting with a neighbor about the weather, maintaining friendships, or making a positive impression on the first day of a new job.

Unlike more structured interactions, which can be formalized to teach more easily as a script [7, 1], small talk demand quick thinking, social flexibility, and the ability to interpret subtle cues such as tone, timing, and context. Training to improve such skills in adulthood presents unique challenges compared to childhood, as many social habits and patterns are already established by the time individuals reach adulthood [15]. These patterns may manifest in the development of strategies to avoid social situations entirely [6]; hence, the skill development that typically occurs through ongoing social interaction may not have been fully realized or practiced. Furthermore, while children have more opportunities for structured social learning and development through school or therapy, adults with ASD may have fewer chances to actively develop these skills [53, 22, 24].

As a result, interventions for adult learners often require more personalized approaches, targeting specific barriers to communication and focusing on building confidence in real-world conversational contexts. Therefore, in this section, we overview the structure and value of small talk, outline formal intervention strategies that may inform the pedagogical design of small talk training, and highlight the potential for robot-assisted social skills training for ASD. This summary of the literature ultimately reveals a gap: while there is a recognized need for small talk skills, there is limited input from adults with ASD themselves regarding the challenges they face in adulthood and few opportunities for targeted training. To address this gap, we conducted a survey to gather insights from adults with ASD (Sec. III).

A. Structure and Value of Small Talk

While the boundaries of conversation types are fluid, "small talk" has a recognized currency in sociolinguistics and communication studies [12]. It refers to informal, light-hearted exchanges focused on building social connections rather than conveying substantial detail, often covering general, non-controversial topics like the weather and hobbies.

Small talk does not have a strict formula, as it is inherently flexible and context-dependent. However, a typical small talk dialogue follows the general sequence of conversation, beginning with a greeting and ending with a closing remark, while emphasizing specific characteristics at each stage [28, 12, 59]:

1) **Greetings and openers**: Initiating the conversation with a greeting or commenting on a shared experience such as the weather or the immediate environment.
2) **General topics**: Discussing non-controversial and general topics such as hobbies, interests, or recent events.
3) **Reciprocity**: Both participants take turns sharing and responding, maintaining a balanced, equitable, and relevant participation in the conversation.
4) **Closure**: The conversation ends with a closing remark, such as indicating appreciation or a future interaction.

For any individual, the characteristics of small talk highlight the subtlety and skill needed to navigate this form of conversation effectively. For adults with ASD, there is a notable overlap between the challenges inherent in small talk and the broader difficulties they report facing in everyday social interactions. We discuss this overlap further in Sec. III.

B. Current Approaches to Small Talk Training

Addressing the unique challenges of small talk for individuals with ASD requires targeted interventions. Although there are currently no programs focused solely on small talk, many broader training initiatives include elements that indirectly support small talk competency. Moreover, while the limited availability of such training for adults with ASD has not been widely explored or critiqued in the literature, the following sections review broader communication programs, which are primarily focused on *children* with ASD. We then examine how these established methods could potentially address the specific challenges of small talk for adults with ASD.

1) Didactic Approaches: Didactic approaches, or classical Applied Behavior Analysis (ABA), break skills into smaller components and train each through highly structured, drill-like practice [50, 17, 15]. While didactic methods have proven effective in various intervention studies for ASD [78, 13], they heavily depend on teacher guidance, prompted responses, and contrived reinforcement methods [50]. An inherent limitation of didactic methods lies in their tendency to encourage passive communication, wherein individuals respond to prompts but may struggle to initiate communication or apply learned behaviors beyond the specific training context.

2) Naturalistic Approaches: Contemporary or naturalistic ABA strives to incorporate interventions into an individual's everyday environment. While these approaches retain some

level of teacher direction, focusing on predetermined goals, they emphasize intrinsic reinforcements, such as personal motivation or social reinforcement. Studies directly comparing didactic and naturalistic approaches have indicated certain advantages of the latter, including better retention and broader application of newly acquired communication skills [14, 50]. *Milieu teaching*, a subset of naturalistic ABA, integrates training into everyday environments to effectively promote spontaneous communication and initiation in individuals with ASD [11, 38, 30, 47]. This method encourages skill development through activities that occur organically throughout the day, rather than being confined to a designated "therapy time."

C. Robot-Assisted Social Skills Training for ASD

There is considerable evidence that technology-driven, practice-based interactions can enhance social skills in adults with ASD [9, 58]. However, robots offer distinct advantages over other technologies by providing a physical, embodied presence that naturally demands a social response [4, 35]. A socially assistive robot (SAR) may feature human-like attributes, such as a face capable of mimicking human expressions or the ability to make eye contact, both of which can elicit social responses from users. This presence allows users to engage in consistent, real-world practice, providing access to repeatable, co-present social interactions that may be challenging to replicate with human therapists [66] or in naturally-occurring, everyday situations.

Several studies have demonstrated the effectiveness of robot-assisted therapy for ASD, with evidence showing improvements in a variety of social and behavioral outcomes. For instance, research has indicated that SARs can foster prosocial behaviors [16], sustain attention [61], elicit spontaneous and appropriate social behavior [67], reduce stereotyped and repetitive behaviors [72], optimize cognitive learning gains [63], and heighten social engagement [66, 52].

While the majority of these studies have focused on children, the growing body of research supports the efficacy of SARs for adults with ASD as well [41, 58]. A robot designed for social skills training could therefore be a valuable tool for adults with ASD, offering a safe, consistent, and adaptive platform for practicing and refining their skills.

III. SURVEY ON THE NEED FOR SMALL TALK TRAINING

To supplement insights from existing literature, we conducted an initial survey to explore the firsthand experiences and challenges faced by adults with ASD. This needs-finding survey aimed to identify specific conversational areas where interventions are not only desired but also deemed most valuable by this population. We also contextualize the survey results within existing research on ASD.

Fifty adults with ASD (22 men and 28 women)[1], ranging from 20 to 55 years ($M = 31.4$, $SD = 10.1$) responded to our

[1]This sample's nearly balanced gender ratio differs from the typical 3:1 male-to-female ratio in ASD diagnoses [37]. This discrepancy may be due to factors such as increased awareness of underdiagnosis in women, recent increases in diagnosis rates, or specific recruitment material or methods in Prolific, which likely attracted a more diverse group of participants.

online survey administered on Prolific [49]. All participants reported having been clinically diagnosed with ASD, either in childhood ($N = 18$) or adulthood ($N = 32$). The survey objectives, design, and analysis were preregistered [60].

A. Small Talk Skills & ASD

While the survey was designed to broadly explore conversational skills in the context of adult social interactions, the majority of respondents emphasized skills closely associated with small talk. When asked in an optional, open-ended prompt about which conversational skills they wished to improve, 48 adults with ASD (96%) specifically expressed a desire to enhance skills that are central to small talk interactions. These reported challenges and desired skills are grouped into five themes as outlined below.

Difficulty Initiating Conversations. Individuals with ASD may find it challenging to initiate conversations, particularly in unfamiliar social situations [51]. Initiating conversations often relies on subtle social cues, such as recognizing when to engage, gauging the appropriate timing for greetings, and responding in a manner that aligns with the context. Among the adults with ASD surveyed, 27 individuals (54%) reported initiating conversations is a skill they wished to improve.

Limited Interest in Non-Specific Topics. Individuals with ASD often exhibit a preference for structured and predictable interactions [54, 51]. The open-ended and non-specific nature of small talk may be discomforting for individuals who prefer environments with well-defined rules and expectations.

Furthermore, individuals with ASD may exhibit a strong interest in specific topics or subjects [44, 69], often preferring in-depth discussions over broad, superficial exchanges. P_{19} stated, "I'm not able to come up with and react to lighthearted banter quickly enough, so I come across as quiet and serious. If I don't know someone well enough, I have no idea what kind of information they would like to receive [...] so I just stay quiet" The preference for depth and detail can make it challenging to engage in the more superficial content typical of small talk. Among those surveyed, 20 individuals (40%) highlighted the ability to discuss general topics beyond their own interests as a specific skill they wished to improve.

Lack of Reciprocal Exchange. Small talk relies on a timely back-and-forth exchange of information and active listening. Individuals with ASD often face difficulties in maintaining reciprocity during conversations, leading to challenges in sustaining the flow of dialogue [44, 46, 75, 32]. Among the adults with ASD surveyed, 39 (78%) expressed a desire to improve their ability to maintain balanced and reciprocal conversations.

Insistence on Topic Sameness. Transitioning smoothly between different topics or concluding the conversation is a skill often required in small talk. Adults with ASD may find it challenging to navigate these transitions, leading to potential disruptions in the flow of conversation [79, 23]. Among the adults with ASD surveyed, 18 individuals (36%) reported that smoothly transitioning between different topics is a valuable conversational skill that they wished to improve.

Social Anxiety. The nuances of social cues can exacerbate feelings of anxiety, leading to the avoidance of social situations altogether [5]. Although reporting wanting to improve in the aforementioned skills, P_{27} shared, "I feel a debilitating consciousness about my eye contact and posture [...] Even if I end up talking, I'm never sure whether it was the right thing." Small talk requires maintaining conversational flow while interpreting nonverbal cues, which can heighten anxiety. On a 7-point Likert scale (1 = highly uncomfortable, 7 = highly comfortable), adults with ASD reported feeling moderately uncomfortable ($M = 2.3$, $SD = 1.4$). 35 participants (70%) identified managing anxiety as a skill they wished to improve.

B. Methods & Challenges to Improvement

Improving small talk skills is seen as highly valuable by adults with ASD, as it plays a critical role in fostering social connections and achieving personal and professional goals. In an optional open-ended response, 40 of the 50 surveyed adults with ASD described the value of this kind of informal conversation in specific aspects of their own lives. We conducted a thematic analysis on their responses, ultimately grouping the descriptions into three primary themes: making new friends (60%), dating (33%), or finding and maintaining a job (60%). Adults with ASD explained that conversations such as "small talk" would be useful to "getting to know people and keeping them interested" (P_{16}), and in "job interviews as [...] you have to use a specific but casual enough talking structure to be considered adequate" (P_{25}) or "socially competent" (P_{48}).

Yet, few adults with ASD ($N = 10$, 20%) reported undergoing formal training in casual conversations through a vocational program, an online class, or coaching from a therapist specialized in interpersonal communication. Despite actively seeking training, the majority of adults (78%) reported having limited or no formal opportunities ($N = 10$) or relying on informal methods ($N = 29$), such as improving their conversational skills through observation or seeking feedback from friends and family.

Generally, humans learn to analyze the interactions they observe and deduce the rules from everyday, naturally-occurring exposure. In free-form responses on the methods that were most beneficial for conversational skill development, 32 adults with ASD (64%) described learning through mere practice and observation beyond or absent of any explicit training. However, natural exposure to an adequate range of real-world interactions, such as the kinds of conversations typical of a workplace, is unlikely for some adults with ASD. To this, P_3 explained that improving in the small talk skills would be valuable for "making friends and going to classes", but "part of the reason I don't do this much is because I don't put myself in situations where I would meet people."

In summary, small talk presents significant challenges for individuals with ASD, particularly in initiating, maintaining, and transitioning within casual conversations. As one participant, P_{37}, explained, "Doing it [small talk] *is* the challenge. Knowing when you should make conversation, knowing the unsaid cues of if a conversation should continue or end.

Knowing what to talk about, not saying too much when asked a question, making sure you ask the next question after. It is a very mentally draining process where I have to evaluate many different factors." In fact, nearly all participants (98%) in this preliminary survey expressed a desire to improve specific skills closely related to small talk. Understanding these challenges is essential for developing and improving access to effective interventions tailored to the unique needs of adults with ASD.

IV. FORMATIVE STUDY

A robot for small talk training should possess proficient small talk ability: balancing conversational succinctness and depth, maintaining an expressive yet appropriate tone, and generating relevant and open-ended responses. Prior research has explored the feasibility of developing robots for small talk interactions [59]. While large language models (LLMs) show substantial potential for enabling natural language capabilities in robots, achieving seamless and contextually appropriate casual dialogue for repeated interactions remains a challenge. Therefore, the first step in developing a robot platform for small talk training is to address this challenge and create a system capable of both generating and evaluating small talk.

Good conversation is believed to arise from the control of low-level attributes [77, 59]. We implemented a grounded observer model—an LLM instance that "observes" ongoing conversations to evaluate whether responses from the "speaking" model adhere to the quantifiable small talk criteria described in Ramnauth et al. (i.e., brevity, tone, coherence, and topic non-specificity) [59]. If the generated response aligns with these criteria, it is relayed; otherwise, the observer generates a revised system prompt and returns it to the speaking model as feedback. This feedback redirection allows the system to self-correct when drifts in conversational behavior are detected. Our implementation is depicted in Fig. 1.

Prior work has not only showed the inadequacy of an "out-of-the-box" LLM for sustaining small talk, but also the observer's robustness in real-time human-robot interactions [59]. Ramnauth et al. [59] found that users were dissatisfied with the baseline LLM's responses, noting its overly assistive and verbose tendencies, which led to conversations described as "rambling" and "like speaking to a wall." In contrast, the observer-enabled system produced conversations that users described as "natural" and "like a casual chat with someone you haven't met before or haven't seen in a long time."

In this present study, we evaluated the robot prototype with the 50 adults with ASD who participated in our needs-finding survey (Sec. III). The goal of presenting this early prototype was to assess whether the observer-enabled small talk reflected relevant real-world conversations and to determine whether adults with ASD would be receptive to both the robot and training. Three randomly selected video excerpts of user interactions were shown. In open-ended feedback, 84% of participants ($N = 42$) reiterated the value of practice-based small talk training for improving confidence (P_{37}), building a habit of engaging in interactions (P_3), and exercising strategies for handling dynamic situations (P_{24}). Participants also noted
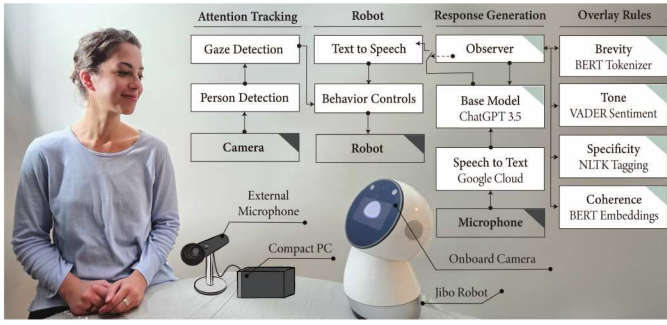
Fig. 1: **Grounded Observer**. Following the framework outlined by Ramnauth et al. [59], we implemented an observer model that enables LLMs to sustain authentic small talk interactions effectively. We then applied this model to a robot, creating a prototype of the training system, and surveyed adults with ASD to gather feedback on whether they would use such a system to improve their small talk skills.

that having the robot at home allows for "practice in a safe environment" (P_{14}) with a "non-judgmental partner" (P_{33}).

Yet, some participants ($N = 12$) expressed wariness to having the robot in their homes. P_{31} noted that, while the robot offers non-judgmental practice, "other people may overhear my conversations." In light of these concerns, several design criteria were established. For example, we aim to design a portable system that operates within a preferred time window, thus allowing users to control when and where interactions take place. Additionally, the system should initiate training only when the user is not engaged in other social activities. The detectors that enable this functionality are described in Sec. VI-B. This design must prioritize user comfort and security, ensuring a personalized and respectful experience. We expand on our design considerations in the following section.

V. DESIGN GOALS FOR ROBOT-ASSISTED TRAINING

Addressing the specific challenges that adults with ASD face calls for targeted small talk interventions. This section lists the core design principles for our robot-assisted training.

A. System Design Objectives

Adhering to the tenets of milieu teaching (Sec. II-C), we designed a robot for small talk training with these objectives:

In a Natural Setting. The system should be tailored for in-home training. This enables users to interact without concern for potential stigma from colleagues. It also eliminates the need for approval or disclosure of a diagnosis to others.

Realistic Interactions. The robot should offer timely and responsive reactions to the user's communication attempts. The robot system must deliver authentic interactions that mirror the appropriateness and style of real-world small talk, responding in real-time and expressing human-like behaviors, including naturalistic gaze, movement, and speech. Additionally, training sessions should not be confined to a designated, scheduled "therapy time," but should occur more organically throughout a portion of the user's day.

User-Led Interactions. The robot's design should empower users to take the lead in interactions by responding to their cues and adjusting its behavior accordingly. A social robot inherently facilitates this objective through its physical presence in the training experience, making it challenging for users to ignore even non-verbal prompts for interaction [4, 19, 58].

Autonomous Behavior. Training must be entirely autonomous, eliminating the need for technical expertise to adjust or control the system once it is given to the user. Although similar systems for ASD have been designed for clinical or laboratory settings where environmental conditions can be controlled or planned for [67], the home is a dynamic, unstructured environment that demands more complex sensing.

B. Training Design Objectives

To design a training method, we break down small talk interactions into smaller, more manageable components and tailor the training method to the unique needs of individuals with ASD. Given the challenges of small talk for individuals with ASD (Sec. III-A), the training focuses on four primary components: initiating a conversation, discussing a non-specific topic, maintaining a flow of dialogue, and appropriately transitioning to a new topic or concluding the interaction.

A tiered training method that models the natural flow of dialogue would sequentially address each component, starting with a simple greeting. The robot responds promptly and contingently, reinforcing positive behaviors. Progressing through the tiers, users tackle more complex aspects, such as introducing non-specific topics or maintaining conversational flow. This approach, illustrated in Fig. 3, encourages users to practice the core components of small talk within each training interaction. Additional considerations include:

Fading Robot-Initiated Prompts. The robot "wakes up" from its idling state to draw attention to the beginning of a training session (Fig. 3-A). In initial interactions with the user, the robot provides an explicit verbal prompt, such as saying, "The training window has started. Remember to make eye contact and greet me." As the user becomes accustomed to the initiation process, the robot gradually minimizes the verbal prompt: "The training window has started." with a less pronounced emphasis. Over time, it is expected that the user initiates to interaction without any specific cue (Fig. 3-B).

Relevant Feedback. The robot should offer clear, supportive feedback, highlighting both user successes and areas for improvement. This encourages users to reflect on their conversations and identify ways to enhance their skills.

Session Length. The optimal amount of training will vary among individuals, but a general guideline is to engage in daily practice. Consistent, short sessions with the robot would be more beneficial than long, less frequent sessions [68, 80].

VI. SYSTEM

This section overviews the hardware and software components that address our design goals (Sec. V) for a fully autonomous, robot-assisted training system for small talk.
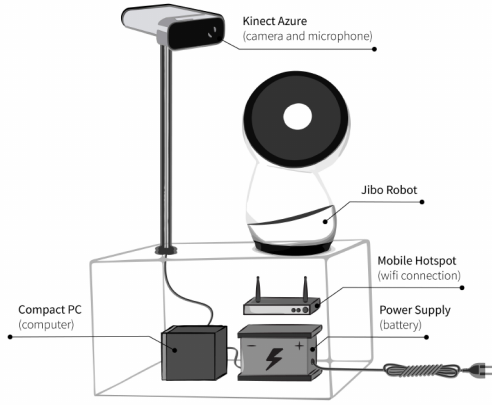
Fig. 2: **System Hardware**. The system has a battery, compact computer, and hotspot that are contained in a hard plastic case. An external camera and microphone are mounted on a mast.

A. Hardware

Our system consisted of five primary hardware components, as shown in Fig. 2. We used the Jibo robot [34], which stands 11 inches tall and features three full-revolute axes for 360-degree movement. Jibo's onboard capabilities enabled naturalistic gaze and movement. A compact PC communicated with other hardware, monitored the system, and served as local data storage during deployment. An Azure Kinect camera was mounted 2 inches above Jibo's head to maximize the visual field and audio capture during training.

The system was designed for self-reliance, equipped with a mobile router that provides a prepaid internet connection for continuous WiFi and automatic cloud-based data synchronization. An uninterruptible power supply served as the charging station. With these components, the setup was plug-and-play, requiring only the charging station to be plugged in. To enhance ergonomic and accessible design, non-interfaceable components were encased in the container on which the robot was placed, reducing apparent complexity.

B. Software

We used a modular software architecture when creating the system to allow for individual components to be evaluated and improved. To achieve this modularity, we created the different components of our software as nodes in the ROS framework [73]. As illustrated in Fig. 1, the system consisted of several components such as attention tracking of the users, robot behavioral control, and the generation of small talk dialogue. These components collectively contributed to small talk interactions described in our initial study (Sec. IV). In addition to these components, we introduced components of training presentation such as the session scheduler and mechanisms for generating relevant feedback to the user.

Delivering the Training Components. A scheduling node determined when the system would begin the training window. One or more training windows could be specified in the system's configurations. During the training window, the system would deliver a session, which consisted of engaging the user

in several small talk conversations. At the end of each session, the system provided feedback, as detailed in Section VI-B.

Recent in-home robot deployments [58, 67, 21] have highlighted the importance of ensuring that in-home systems are capable of discerning when it is socially appropriate to engage users. These studies have emphasized that, for effective interaction, the system must recognize contextual cues that indicate whether the user is in a setting conducive to engagement or learning. Therefore, we incorporate two classifiers: audio-based social presence classification [21] and person detection. The Azure Kinect captured video recordings, and image snapshots were analyzed by a pre-trained YOLO neural network [18] to estimate the number of people present. The Kinect also transcribed audio using Google's Speech-to-Text API to identify speech content. If speech was detected, the social presence classifier determined whether it represented a physically co-present conversation (e.g., a dinner party or a conversation between friends) or a media interaction (e.g., watching television or listening to the radio) [21]. If two or more people were detected and the audio was classified as non-media, the system assumed it was not an appropriate time to engage the user and skipped the planned interaction. However, the frequency of planned conversations increased gradually to maintain a consistent number of interactions within the training window, with intervals selected from a Gaussian distribution to avoid predictability.

When the system was not engaging in a small talk conversation, Jibo performed an idling behavior, ranging from sleeping to looking down at the floor. When prompted to "wake up", Jibo looked up at the user to signal its availability to chat. The user could initiate conversation with a greeting, or Jibo would do so after some time. A linear-logarithmic growth function determined how long Jibo waited for user initiation, progressively extending this window. If the user does not respond, Jibo would prompt again.

Following a greeting, the system randomly selected 8 to 12 rounds of conversation. It used an observer model (Sec. IV) to ensure compliance with small talk criteria and monitor user progress. After the selected rounds are completed, Jibo transitioned to feedback mode, signaled by changing its LED ring to blue and adopting a more formal voice.

Generating Feedback. Feedback is essential in any training. In their study, Ramnauth et al. [59] found that quantitative definitions of small talk criteria (e.g., brevity, tone, non-specificity, coherence) effectively informed an observer model for generating feedback to the speaking model. However, these quantitative definitions do not easily translate into practical feedback for human users.

Effective user feedback should be specific, timely, constructive, objective, and respectful. We used a separate instance of GPT-3.5 that has a prompt delineating these characteristics of good micro- and macro-level feedback. *Micro-level feedback* highlighted user successes after each conversation, along with one area for improvement. For instance, "You asked great questions about my favorite hobbies. I would like to hear about your hobbies next time." *Macro-feedback* summarized
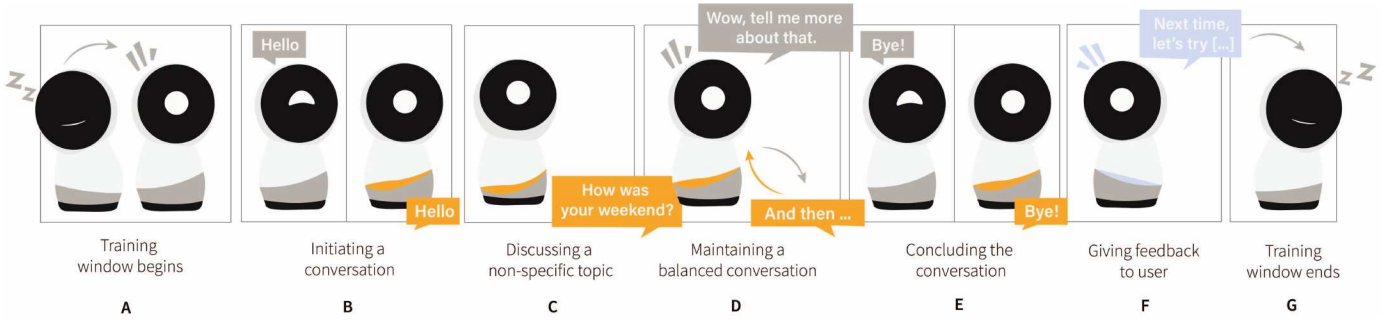
Fig. 3: **Training Sequence**: A training session unfolds in distinct stages: (A) The robot "wakes up" to signal the start of the session, (B) encouraging users to initiate with a greeting. Users practice small talk skills, including (C) discussing non-specific topics, (D) maintaining balanced dialogues, and appropriate transitions (E). Users complete several conversations, with the robot giving micro-level feedback after each conversation (B-F). At the end of session, the robot gives the user macro-level, overall feedback (F). Outside of the specified training window, the robot performs an idling behavior (G).

all previous conversations within a training window, providing overarching insights. For example, "You did a great job acknowledging my interests; I enjoyed your rainy day activity suggestions. For our next conversations, try discussing your weekend plans or the weather." Additionally, if a small talk rule was consistently or severely broken during a training session, the robot would act out a dialogue to demonstrate following that rule. In these demonstrations, Jibo modulated its speech pitch and duration and changed its LED color. It even physically turned to "face" itself as it alternated between characters, like a puppet speaking to itself. Altogether, we aimed to explore if this generated feedback was effective, constructive, and well-received by adults with ASD.

Robustness for Contactless, Home Deployments. Robots deployed in homes require much greater robustness than those in controlled lab settings. The unstructured home environment poses challenges like power outages, fluctuating lighting, and unexpected user distractions. To enhance our system's reliability, we implemented watchdog scripts to monitor performance and remote desktop applications for troubleshooting. One script ran at the start of each training session to verify the camera, microphone, and communication with Jibo. The second script ran after each day to check the file sizes of recordings. The scripts would send an email detailing component success or failure. Remote access allowed for remote configuration; the system could be delivered to the user's home and then configured completely without in-person contact.

VII. IN-HOME DEPLOYMENTS

The study was preregistered and approved by the university's Institutional Review Board. Interested adults with ASD enrolled through the study's website promoted via channels that required a clinical diagnosis of ASD, such as ASD-specific residential facilities or employment networks. To comply with COVID-19 safety protocols, the system was designed for easy, contactless setup. It was delivered directly to participants' homes with detailed written and video instructions. Remote support was available via phone or video call, and at no point did a researcher enter participants' homes.

Participants placed the system in a comfortable room and specified a daily training window of up to three hours. The study lasted at least seven days, though adjustments were made for scheduling conflicts. After the study, participants took part in an interview to discuss their experience with the system, its effectiveness, and suggestions for improvement.

A. Data Collection

Video and audio recordings of all training interactions captured participants' responses to the robot. It is well-known that automated systems, particularly those relying on computer vision or speech recognition, often face challenges with environmental noise or the diverse behaviors exhibited by participants, especially those with ASD [61]. As a result, we opted to manually annotate the collected dataset and verify the accuracy of inputs, such as speech transcripts, before applying automated measures. Using ELAN [40], three undergraduate research assistants annotated the start and stop times for each robot and user response. These markers allowed us to calculate users' *initiation rate* defined as the proportion of conversations initiated by the user to all conversations within a session, and the duration of each *conversational turn* which is the time one speaker (robot or user) spent talking before the other responded. Additionally, for each turn, a binary label indicated whether the user made eye contact with the robot.

To ensure the reliability of these annotations, inter-rater reliability scores were calculated for each of the annotated metrics. Cohen's Kappa (κ) was used for categorical variables, such as eye contact, and the intraclass correlation coefficient (ICC) for continuous variables, such as initiation rate and conversational turn duration. The resulting κ for eye contact was 0.95, indicating a high level of agreement between the annotators in identifying whether the user made eye contact with the robot during a given conversational turn. For the continuous metrics, the average ICC for initiation rate was 0.91, and for conversational turn duration, it was 0.93. These scores demonstrate strong consistency among annotators in evaluating the training interactions.

Furthermore, transcripts generated by Google's Speech-to-Text were manually reviewed and corrected, particularly for

Fig. 4: **In-Home Deployments**. The collage on the left shows in-home interactions with adults with ASD from the system's point-of-view. On the right is the system placed in a user's living room after the contactless delivery.

users with accents or atypical intonations. Each response was assessed using the observer's evaluative metrics for small talk—brevity, tone, specificity, and coherence—as defined by Ramnauth et al. [59].

These annotations and observer-derived metrics enabled us explore behavioral trends over the course of the study, focusing on the frequency of user eye contact, initiative, small talk violations, and overall conversational dynamics. This quantitative analysis was further enriched by qualitative insights from post-study interviews, which explored participants' comfort with the robot, their perception of their own conversational skills, and the relevance of the robot's training and feedback.

B. Participant Information

Twenty five adults with ASD (19 males, 6 females), ranging from 18 to 68 years ($M = 32.4$, $SD = 12.7$) participated in the study. All had a confirmed diagnosis of ASD and completed the Autism Spectrum Quotient-10 (AQ-10) survey prior to the study. Fifteen participants were classified as high-functioning, with an average AQ-10 score of 4.8 ($SD = 1.2$). The remaining ten had higher scores ($M = 7.1$, $SD = 2.4$) and lived with caregivers. In this group, four participants were also diagnosed with Intellectual Disability (ID), four with Attention Deficit Hyperactivity Disorder (ADHD), two with Obsessive-Compulsive Disorder (OCD), and one with Down syndrome. All with co-occurring conditions were receiving medication and specialized therapy or education at the time of this study.

C. Results

A total of 2,114 conversations were recorded, yielding 9,870 user responses across 225 sessions or 281.3 hours of participant interaction. Participants experienced an average of 8.9 sessions ($SD = 2.3$) over the course of the study.

Initiating Conversations. We analyzed the frequency of user-initiated conversations across all sessions to understand the degree of proactive engagement with the system. On average, participants initiated conversations in 34% of interactions during their first session, rising to 64% by the final session. Users with higher AQ-10 scores started with a lower initiation rate of 14% in the initial session, which increased to 55% by the end of their study ($t = 18.9$, $p \leq 0.0001$). Conversely, high-functioning users showed an initiation rate of 50% in the initial session, which increased to 70% by the study's conclusion (t

$= 12.4$, $p \leq 0.0001$). All 25 participants demonstrated some improvement in their initiation rate between the initial and final sessions of the study ($\Delta M = 36\%$, $\Delta SD = 20\%$).

A linear regression on the initiation rate over sessions yielded a significant increase ($\beta = 0.07$, $p \leq 0.0001$). This growth suggests that participants, irrespective of their initial engagement level, became more active in initiating small talk.

Maintaining Eye Contact. We assessed the effect of the training on users' ability to maintain appropriate eye contact. A linear regression analysis predicting eye contact frequency per conversational turn based on session number revealed a significant increase in participants' eye contact with the system over the course of the study ($\beta = 0.02$, $p \leq 0.001$). On average, participants maintained eye contact in 24% of turns during the first session, compared to 51% in the final session. Additionally, all 25 participants showed improvement in eye contact frequency from the initial to final session, with an average increase of 32% ($\Delta SD = 8\%$).

Sustaining Appropriate Small Talk. We chose to separate the greeting phase from the rest of the conversation because greetings, such as "hello" or "good morning," tend to be routine, shorter and less variable. In contrast, post-greeting responses are more varied, context-dependent, and reflect the true conversational dynamics. Greetings were identified through keyword matching and checking if the sequential response index is within the first full conversational turn. Post-greeting user responses ($N = 6,751$) were assessed using the observer's metrics and behavioral annotations.

Brevity. The average turn duration of user responses in seconds (s) exhibited a significant increase by session ($\beta = 0.77$, $p \leq 0.0001$). The average turn duration was $4.4s$ ($SD = 1.5$) in the first session as compared to an average duration of $5.3s$ ($SD = 2.8$) in their final session. This is supplemented by the observer brevity metric based on word count; the observer flagged 2,421 user responses (21%) as being overly verbose, and produced macro-level feedback on keeping responses more concise 61 times, or 27% of all sessions.

Further analysis revealed a significant increase in the robot's turn duration ($\beta = 0.79$, $p \leq 0.0001$), indicating that its responses lengthened over time. Users' conversational balance, defined as the proportion of turn duration between their previous utterance and current response, also improved significantly ($\beta = 0.10$, $p \leq 0.0001$). This suggests that users provided longer responses and engaged in more balanced, reciprocal interactions as the study progressed.

Non-specificity. The observer-derived metrics for non-specificity evaluate the frequency of named entities and descriptive words in interactions. Due to the nature of ASD and the associated challenges with theory of mind, it is common for individuals with ASD to mention named entities (e.g., a cat named Cheddar or a street called Cedar) without considering that the robot lacks the contextual knowledge needed to respond appropriately. In the post-study interview, the mother of P_{19} recounted, "Kurt often got confused, sometimes frustrated, when the robot misunderstood him. For example, when he mentioned feeding his cat, Cheddar, the robot started asking

about cheeses." Eight other users, all with high AQ-10 scores, demonstrated similar difficulties during their interactions.

A linear regression on named entity counts showed a significant increase by session ($\beta = 0.01$, $p = 0.002$), as did descriptive words ($\beta = 0.03$, $p \leq 0.0001$). Notably, only 16 participants exhibited these increases, with all but one categorized as high-functioning based on AQ-10 scores. The observer flagged 1,223 user responses (12%) as overly specific, providing macro-level feedback on broadening responses in 83 sessions (37% of total). Importantly, 85% of these violations occurred after the first three sessions, indicating that users, particularly higher-functioning adults, felt more comfortable engaging in deeper conversations as the study progressed.

Tone. There was no significant change in tone during the study ($\beta = 0.00$, $p = 0.69$). The observer metric for tone based on sentiment analysis yielded macro-level feedback on keeping responses more positive 27 times (12% of all sessions).

Coherence. There was no significant change in user's response coherence ($\beta = 0.00$, $p = 0.82$). The observer's coherence metric based on information gain yielded feedback on keeping responses relevant 54 times (24% of all sessions).

System Performance. A key indicator of the system's performance is sustained user engagement throughout the study. Above, we detailed trends in the frequency of user-initiated interactions and levels of eye contact with the robot. These are simultaneously considered measures of system performance and user improvement as a result of robot-assisted training.

Moreover, while Ramnauth et al. [59] showcased the effectiveness of the grounded observer in novel, in-person user interactions, we examine its performance in this context of in-home, long-term small talk training for individuals with ASD. Of all robot-delivered responses to the user ($N = 9,870$), the instances in which the observer instructed the speaking model to regenerate its proposed response ($N = 8,585$) offer valuable insights into the interaction dynamics and the effectiveness of the observer's feedback loop. In our system design, we limited the feedback loop to a maximum of three regenerations to ensure the robot delivered a timely response. From this, the first regeneration revealed the base LLM's natural tendencies, the second highlighted any gaps in the observer's feedback, and the third indicated potential issues either reintroduced or overlooked by that feedback. In our study, 65% ($N = 6,415$) of initially proposed responses by the base model required regeneration, with brevity (45%) and non-specificity (19%) being the most common reasons. This indicates a tendency of the LLM to produce overly verbose and detailed responses that needed refinement. The second regeneration, which occurred in 21% ($N = 2,072$) of these cases, primarily addressed coherence (11%) and brevity (6%), suggesting the observer's initial feedback in these cases resulted in less coherent, excessively terse corrections. Finally, only 1% ($N = 98$) of proposed responses required a third regeneration, with brevity being the only recurring issue.

Since the efficacy of system prompts is often assessed through post-hoc performance or trial-and-error testing, it is challenging to quantify the exact improvement that can be achieved through a more crafted initial or feedback prompts. Nevertheless, these results demonstrate that the observer was largely successful in realigning the base LLM's deviations from appropriate small talk conventions.

VIII. DISCUSSION

Here, we reflect on our findings, incorporating insights gained from participant interviews. Overall, the study showed clear improvements in users' conversational skills and engagement with the training system. As training progressed, users showed greater confidence and ease in initiating conversations, maintaining eye contact, and engaging in more balanced, yet detailed conversations over time. Interestingly, while longer, more descriptive responses were flagged as violations of small talk norms—resulting in the robot giving more frequent feedback to maintain brevity—these deviations suggest an increasing preference for more substantial conversations over time. Together, these results reflect users' growing sense of ease and interest in the robot as a conversational partner.

A. Post-Study Interviews

Interviews were conducted with 24 participants with ASD[2] and 12 primary caregivers, 10 of whom lived with a participant during the study. These semi-structured post-study interviews, conducted by a member of the research team, lasted between 30 and 45 minutes. The interviews were recorded and transcribed to ensure accuracy and participant confidentiality. Caregivers were invited to join the interview only with the prior consent of the participants. Based on the feedback collected, we performed an informal thematic analysis, categorizing the insights from participants and caregivers into three main themes, as outlined below.

Engagement with Training. Participants widely reported that the robot's structured yet dynamic interaction style helped make small talk practice approachable. Several participants mentioned that its unpredictable behavior (e.g., random sleep intervals and spontaneous prompts) mimicked real-life conversations, which kept interactions "fresh and challenging" (P_3). This aligns with previous findings that highlight the importance of unpredictability in training, as it more closely simulates natural, unscripted social encounters [14, 50].

Most conversations adhered to typical small talk topics (84% of all user responses), though some users engaged in deeper discussions, including personal experiences like bullying at work or the loss of a family member. While some users ($N = 9$) felt they had "quickly exhausted the limited range of topics considered to be small talk" (P_3), others ($N = 12$) appreciated that even simple prompts about their day were "meditative, and made me reflect on the day and my feelings in a more mindful way," (P_2). Several users noted that, beyond skills training, the small talk robot provided valuable mental and emotional support. P_{20} shared, "My daily life can be isolating, so I don't often get asked how my day is going or how I'm doing [...] I enjoyed these kinds of chats when I'm

[2]One individual was unable to participate in the post-study interview due to unforeseen personal circumstances.

making breakfast or coming home from a long day, watching TV, and there's no one else to talk to."

The robot also served as a valuable medium for facilitating communication between caregivers and their child. Several parents ($N = 5$) noted that the robot's "neutral, non-judgmental presence" (P_{12}) made it easier to initiate conversations on difficult topics. The parent of P_{14} shared, "We've had trouble talking about certain feelings, but when the robot asked about how his day was going, it opened up a way for us to talk. It felt like a safe space for him to express what he normally wouldn't share with me directly."

Perceived Skill Improvement. Adults with ASD ($N = 18$) and their caregivers ($N = 10$) remarked on behavioral gains which were not captured in the observer's metrics. The robot's training was designed to simulate real conversations, complete with pauses, turns, and varied topics, which helped participants grow more confident and adept at holding discussions. For example, P_{11}'s parent reflected, "My son has made great progress in understanding conversational turn-taking. He used to dominate conversations or struggle with waiting for his turn, but the training has helped him better grasp the flow of dialogue." Another parent observed how P_{17} became more confident in her interactions with others: "It's been incredible to see my daughter take more initiative in our family interactions. She's been practicing with Jibo, and now she actively participates in conversations at the dinner table, offering her own thoughts and even asking others about their experiences."

Relevance of Robot Feedback. Although many robot-assisted social skills interventions for individuals with ASD exist [52], most are designed primarily for children [58] and focus on practice rather than direct feedback. This is largely because social skills are personal and highly individualized [71], making it difficult to provide constructive feedback that resonates with each user. However, giving direct feedback is a core component of our training pedagogy. This approach to feedback could have broader implications for how we design robot-directed interventions for adult populations—while children may benefit from simpler forms of reinforcement, the robot's ability to provide explicit, personalized feedback was seen as highly valuable by many of our participants ($N = 21$).

This approach not only supported skills improvement but also encouraged cognizance of how one interacts with others. P_{10} shared, "During one session, Jibo and I discussed our favorite foods. The conversation lasted longer than usual because Jibo encouraged me to ask follow-up questions. By the end, Jibo gave me tips and even an example on how to keep conversations engaging—advice I've since used with my coworkers." Many users ($N = 13$) commented on being more mindful and reflective about the quality of their conversations as a result of the robot's feedback. P_{14} explained, "It was new to me, practicing this kind of mindfulness—thinking more deeply about how and what to say."

B. Ethical Considerations

In general, the deployment of LLM-based systems in personal settings for vulnerable users raises ethical concerns. For example, as motivation for developing the grounded observer framework, Ramnauth et al. described how inaccurate or misaligned responses could pose safety-critical risks [59]. Furthermore, the system must prioritize user privacy and security to prevent misuse of sensitive personal data.

At the time of this study, local use of GPT-3.5 and many other LLMs were not supported. While open-source alternatives offered similar architectures, they required extensive computational resources, making them impractical for real-time interactions that rely only on the robot's on-board capabilities and a compact PC. Other stable models, including GPT-2 [57] and LLAMA 2 [74], were trained on smaller datasets, resulting in less mature natural language capabilities and an increased risk of generating harmful language.

To further mitigate these risks, data was transmitted to the cloud only during active participant engagement in training sessions, minimizing the amount of participant data sent throughout the in-home deployment. While our informed consent process provided participants with a clear and accessible explanation of the cloud-based LLM's use, we encourage researchers and developers to thoughtfully balance a model's readiness for real-time interaction with considerations of privacy, security, and computational feasibility.

C. Study Limitations & Directions for Future Work

It is important to note that the present study was conceived as an exploratory investigation rather than a hypothesis-driven evaluation. Our primary objective was to demonstrate the *feasibility* and process of designing, developing and deploying a social robot platform for long-term, in-home interactions with an understudied population. Still, lasting behavioral changes from training typically require months or years to manifest. Although our approximately week-long intervention captured early engagement patterns and potential novelty effects, longer-term studies are needed to assess the sustainability of any observed improvements.

Further, given the practical challenges of establishing a comparable control condition in naturalistic settings, we adopted a within-subject design in which each participant served as their own baseline—a common approach in ASD research due to high inter-individual variability. We acknowledge the limitations of this method and refrain from making strong claims of efficacy. For instance, one may argue that increases in social initiations and eye contact could reflect familiarity or novelty effects. Interestingly, one may even expect the opposite—frequently repeated interactions, especially with a system of limited personalization or interaction memory, would result in *decreased* engagement over time. While we cannot attribute user outcomes solely to the system's design, the system shows promising potential in facilitating *continued* engagement with the robot-led training. Nonetheless, larger-scale studies with extended training periods will be essential for assessing the *efficacy* and long-term impact of this approach.

Moreover, this study was conducted within a single country and cultural context, which may limit the generalizability of its findings to other regions and populations. Differences in

communication styles, social norms, approaches to the diagnosis or treatment of ASD, and attitudes toward technology across cultures can significantly influence both the acceptance of social robots and the effectiveness of the intervention [36]. To gain a more comprehensive understanding of the feasibility and efficacy of robot-assisted training, future research should replicate and expand upon this work in diverse cultural and geographic settings, accounting for the unique ways these factors may shape user engagement and outcomes.

Another limitation is that the behavioral detection and response generation of our robot system rely on models like Google's Speech-to-Text for speech processing and OpenFace for gaze and pose estimation, which are largely trained on data from neurotypical individuals in controlled environments. Consequently, deploying such models in in-home settings with adults with ASD introduces challenges due to the unique characteristics of this population and the unpredictable nature of real-world environments [61]. For instance, some participants in our study exhibited atypical speech disfluencies, such as irregular pauses, repetition, or unexpected intonation patterns, which are common in individuals with ASD [65]. These patterns can cause the system to misinterpret or entirely fail to recognize user inputs, leading to breakdowns in conversational flow. Similarly, inaccuracies in gaze estimation can result in the system failing to distinguish between the user and other visual stimuli in the environment, such as faces appearing on a TV or reflections in mirrors. This occasionally caused the robot to direct its gaze away from the user, potentially reducing the perceived quality and engagement of the interaction. Although these issues were not mentioned in our post-study interviews, we observed them while reviewing the dataset to correct errors in the generated speech transcripts. Given the growing interest in automatic behavioral annotation [76], our observations underscore the need for model adaptations or fine-tuning to better align with the needs of diverse populations and naturalistic settings.

IX. CONCLUSION

Engaging in small talk is a vital social skill that impacts everyday interactions, from making friends to job interviews. For adults with ASD, small talk poses unique challenges yet is essential for social integration and professional opportunities. In this study, we developed an autonomous, robot-assisted training platform to enhance small talk skills. While significant, long-lasting behavioral change typically require weeks or months of training, our results showed that even within a week, adults with ASD made meaningful progress, showing increased initiative in conversations and improved eye contact. This study makes several key contributions to our understanding of robot-assisted skills training: (1) we demonstrate how relevant training can be designed to target desired skills for an understudied population, as the majority of ASD research focuses on children and very little on adults; (2) we demonstrate the feasibility of creating a fully autonomous, plug-and-play system for daily in-home practice; (3) we emphasize the value of personalized, direct feedback beyond

mere practice in social skills training; and (4) we underscore the value of an otherwise overlooked form of communication (small talk) not just for social skills development, but also for supporting positive mental and emotional outcomes for users.

REFERENCES

[1] Jessica S Akers, Nicole Pyle, Thomas S Higbee, Daniel Pyle, and Kristina R Gerencser. A synthesis of script fading effects with individuals with autism spectrum disorder: A 20-year review. *Review Journal of Autism and Developmental Disorders*, 3:1–17, 2016.

[2] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[3] Franziska Babel, Johannes Kraus, Linda Miller, Matthias Kraus, Nicolas Wagner, Wolfgang Minker, and Martin Baumann. Small talk with a robot? the impact of dialog content, talk initiative, and gaze behavior of a social robot on trust, acceptance, and proximity. *International Journal of Social Robotics*, pages 1–14, 2021.

[4] Wilma A Bainbridge, Justin Hart, Elizabeth S Kim, and Brian Scassellati. The effect of presence on human-robot interaction. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 701–706. IEEE, 2008.

[5] Lindsey Balderaz. Social skills interventions for adults with asd: a review of the literature. *Journal of Psychosocial Rehabilitation and Mental Health*, 7(1):45–54, 2020.

[6] Susanne Bejerot, Jonna M Eriksson, and Ewa Mörtberg. Social anxiety in adult autism spectrum disorder. *Psychiatry research*, 220(1-2):705–707, 2014.

[7] Scott Bellini and Jessica K Peters. Social skills training for youth with autism spectrum disorders. *Child and adolescent psychiatric clinics of North America*, 17(4): 857–873, 2008.

[8] Simon Bristoll and Jules Dickinson. Small talk, big results. *Newsli*, 92:6–13, 2015.

[9] Moira Burke, Robert Kraut, and Diane Williams. Social use of computer-mediated communication by adults on the autism spectrum. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 425–434, 2010.

[10] John-John Cabibihan, Hifza Javed, Marcelo Ang, and Sharifah Mariam Aljunied. Why robots? a survey on the roles and benefits of social robots in the therapy of children with autism. *International journal of social robotics*, 5:593–618, 2013.

[11] Robyn J Christensen-Sandfort and Stacie B Whinnery. Impact of milieu teaching on communication skills of young children with autism spectrum disorder. *Topics in Early Childhood Special Education*, 32(4):211–222, 2013.

[12] Justine Coupland. *Small talk*. Routledge, 2014.

[13] M Dehqonova and G Tagonova. The importance of didactic games in speech therapy in the development of speech in children with autism and the ability to choose effective methods. *European Journal of Innovation In Nonformal Education*, 2(1):133–137, 2022.

[14] Dennis J Delprato. Comparisons of discrete-trial and normalized behavioral language intervention for young children with autism. *Journal of autism and developmental disorders*, 31:315–325, 2001.

[15] Francis J DeMatteo, Patricia S Arter, Christie Sworen-Parise, Michael Fasciana, and Marcie A Paulhamus. Social skills training for young adults with autism spectrum disorder: Overview and implications for practice. *National Teacher Education Journal*, 5(4), 2012.

[16] Joshua J Diehl, Lauren M Schmitt, Michael Villano, and Charles R Crowell. The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1): 249–262, 2012.

[17] Chetna Duggal, Bakul Dua, Ritika Chokhani, and Koyeli Sengupta. What works and how: Adult learner perspectives on an autism intervention training program in india. *Autism*, 24(1):246–257, 2020.

[18] Ali Farhadi and Joseph Redmon. Yolov3: An incremental improvement. In *Computer Vision and Pattern Recognition*, pages 1804–02767, 2018.

[19] Stephen M Fiore, Travis J Wiltshire, Emilio JC Lobato, Florian G Jentsch, Wesley H Huang, and Benjamin Axelrod. Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior. *Frontiers in psychology*, 4:859, 2013.

[20] Veerle Garrels. Getting good at small talk: Student-directed learning of social conversation skills. *European Journal of Special Needs Education*, 34(3):393–402, 2019.

[21] Nicholas C Georgiou, Rebecca Ramnauth, Emmanuel Adeniran, Michael Lee, Lila Selin, and Brian Scassellati. Is someone there or is that the tv? detecting social presence using sound. *ACM Transactions on Human-Robot Interaction*, 12(4):1–33, 2023.

[22] Peter F Gerhardt and Ilene Lainer. Addressing the needs of adolescents and adults with autism: A crisis on the horizon. *Journal of Contemporary Psychotherapy*, 41: 37–45, 2011.

[23] Hilde M Geurts, Blythe Corbett, and Marjorie Solomon. The paradox of cognitive flexibility in autism. *Trends in cognitive sciences*, 13(2):74–82, 2009.

[24] Janet E Graetz. Autism grows up: Opportunities for adults with autism. *Disability & Society*, 25(1):33–47, 2010.

[25] Janet Holmes. When small talk is a big deal: Sociolinguistic challenges in the workplace. *Second language needs analysis*, 344:371, 2005.

[26] Karen Hurlbutt and Lynne Chalmers. Employment and adults with asperger syndrome. *Focus on autism and other developmental disabilities*, 19(4):215–222, 2004.

[27] Naoki Ise and Takamasa Iio. Social robot encouraging two strangers to talk with each other for their relationships. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 144–147, 2021.

[28] Adam Jaworski. Silence and small talk. In *Small talk*, pages 110–132. Routledge, 2014.

[29] Brianna Joseph, Kelly B Kearney, Michael P Brady, Angelica Downey, and Ayse Torres. Teaching small talk: Increasing on-topic conversational exchanges in college students with intellectual and developmental disabilities using remote audio coaching. *Behavior modification*, 45 (2):251–271, 2021.

[30] Ann P Kaiser, Terry B Hancock, and Jennifer P Nietfeld. The effects of parent-implemented enhanced milieu teaching on the social communication of children who have autism. *Early Education and Development*, 11(4): 423–446, 2000.

[31] Junko Kanero, Vasfiye Geçkin, Cansu Oranç, Ezgi Mamus, Aylin C Küntay, and Tilbe Göksun. Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, 12(3):146–151, 2018.

[32] Lynn Kern Koegel, Mi N Park, and Robert L Koegel. Using self-management to improve the reciprocal social conversation of children with autism spectrum disorder. *Journal of autism and developmental disorders*, 44:1055–1063, 2014.

[33] Hideki Kozima, Cocoro Nakagawa, and Yuriko Yasuda. Interactive robots for communication-care: A case-study in autism therapy. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 341–346. IEEE, 2005.

[34] MIT Media Lab. Jibo social robotic research platform, 2021. URL https://www.media.mit.edu/projects/jibo-research-platform/overview/.

[35] Dan Leyzberg, Eleanor Avrunin, Jenny Liu, and Brian Scassellati. Robots that express emotion elicit better human teaching. In *Proceedings of the 6th International Conference on Human-robot Interaction*, pages 347–354, 2011.

[36] Velvetina Lim, Maki Rooksby, and Emily S Cross. Social robots on a global stage: establishing a role for culture during human–robot interaction. *International Journal of Social Robotics*, 13(6):1307–1333, 2021.

[37] Rachel Loomes, Laura Hull, and William Polmear Locke Mandy. What is the male-to-female ratio in autism spectrum disorder? a systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 56(6):466–474, 2017.

[38] G Richmond Mancil. Milieu therapy as a communication intervention: A review of the literature related to children with autism spectrum disorder. *Education and Training in Developmental Disabilities*, pages 105–117, 2009.

[39] Maja J Matarić and Brian Scassellati. Socially assistive robotics. *Springer handbook of robotics*, pages 1973–1994, 2016.

[40] Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. Elan. URL https://archive.mpi.nl/tla/elan.

[41] Peter E McKenna, Ingo Keller, Jose L Part, Mei Yii Lim, Ruth Aylett, Frank Broz, and Gnanathusharan Rajendran. "sorry to disturb you" autism and robot interruptions. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 360–362, 2020.

[42] Eve Müller, Adriana Schuler, Barbara A Burton, and Gregory B Yates. Meeting the vocational support needs of individuals with asperger syndrome and other autism spectrum disabilities. *Journal of Vocational Rehabilitation*, 18(3):163–175, 2003.

[43] Jacqueline Nadel, Ouriel Grynszpan, and Jean-Claude Martin. Autism and socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, pages 437–462. 2022.

[44] Aparna Nadig, Iris Lee, Leher Singh, Kyle Bosshart, and Sally Ozonoff. How does the topic of conversation affect verbal exchange and eye gaze? a comparison between typical development and high-functioning autism. *Neuropsychologia*, 48(9):2730–2739, 2010.

[45] Eric Nichols, Sarah Rose Siskind, Levko Ivanchuk, Guillermo Pérez, Waki Kamino, Selma Šabanović, and Randy Gomez. Hey haru, let's be friends! using the tiers of friendship to build rapport through small talk with the tabletop robot haru. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6101–6108. IEEE, 2022.

[46] Keiko Ochi, Nobutaka Ono, Keiho Owada, Masaki Kojima, Miho Kuroda, Shigeki Sagayama, and Hidenori Yamasue. Quantification of speech and synchrony in the conversation of adults with autism spectrum disorder. *PloS one*, 14(12):e0225377, 2019.

[47] Billy T. Ogletree, Johanna R. Price, and Jonathan M. Campbell. A selective review of milieu interventions for adults with severe intellectual disabilities. *Current Developmental Disorders Reports*, 7:109–115, 2020.

[48] Alisha Ohl, Mira Grice Sheff, Sarah Small, Jamie Nguyen, Kelly Paskor, and Aliza Zanjirian. Predictors of employment status among adults with autism spectrum disorder. *Work*, 56(2):345–355, 2017.

[49] Stefan Palan and Christian Schitter. Prolific. a subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27, 2018.

[50] Rhea Paul. Interventions to improve communication in autism. *Child and adolescent psychiatric clinics of North America*, 17(4):835–856, 2008.

[51] Rhea Paul, Stephanie Miles Orlovski, Hillary Chuba Marcinko, and Fred Volkmar. Conversational behaviors in youth with high-functioning asd and asperger syndrome. *Journal of autism and developmental disorders*, 39:115–125, 2009.

[52] Paola Pennisi, Alessandro Tonacci, Gennaro Tartarisco, Lucia Billeci, Liliana Ruta, Sebastiano Gangemi, and Giovanni Pioggia. Autism and social robotics: A systematic review. *Autism Research*, 9(2):165–183, 2016.

[53] Elizabeth A Perkins and Karen A Berkman. Into the unknown: Aging with autism spectrum disorders. *American journal on intellectual and developmental disabilities*, 117(6):478–496, 2012.

[54] Valentina Petrolini, Marta Jorba, and Agustín Vicente. What does it take to be rigid? reflections on the notion of rigidity in autism. *Frontiers in Psychiatry*, 14:1072362, 2023.

[55] Penny Pexman, Lorraine Reggin, and Kate Lee. Addressing the challenge of verbal irony: Getting serious about sarcasm training. *Languages*, 4(2):23, 2019.

[56] Patricia Pullin. Small talk, rapport, and international communicative competence: Lessons to learn from belf. *The Journal of Business Communication (1973)*, 47(4):455–476, 2010.

[57] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[58] Rebecca Ramnauth, Emmanuel Adéníran, Timothy Adamson, Michal A Lewkowicz, Rohit Giridharan, Caroline Reiner, and Brian Scassellati. A social robot for improving interruptions tolerance and employability in adults with asd. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 4–13. IEEE, 2022.

[59] Rebecca Ramnauth, Dražen Brščić, and Brian Scassellati. A grounded observer framework for establishing guardrails for foundation models in socially sensitive domains. *arXiv preprint arXiv:2412.18639*, 2024.

[60] Rebecca Ramnauth, Drazen Brscic, and Brian Scassellati. A robot-assisted approach to small-talk training for adults with asd, Jun 2024. URL osf.io/kad27.

[61] Rebecca Ramnauth, Frederick Shic, and Brian Scassellati. Gaze behavior during a long-term, in-home, social robot intervention for children with asd. In *Proceedings of the 2025 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '25, page 949–957. IEEE Press, 2025.

[62] Ben Robins, Kerstin Dautenhahn, and Paul Dickerson. From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *2009 second international conferences on advances in computer-human*

interactions, pages 205–211. IEEE, 2009.

[63] Ben Robins, Kerstin Dautenhahn, and Paul Dickerson. Embodiment and cognitive learning–can a humanoid robot help children with autism to learn about tactile social behaviour? In *International Conference on Social Robotics*, pages 66–75. Springer, 2012.

[64] Anne M Roux, Paul T Shattuck, Benjamin P Cooper, Kristy A Anderson, Mary Wagner, and Sarah C Narendorf. Postsecondary employment experiences among young adults with an autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(9):931–939, 2013.

[65] Kathleen Scaler Scott, John A Tetnowski, James R Flaitz, and J Scott Yaruss. Preliminary study of disfluency in school-aged children with autism. *International Journal of Language & Communication Disorders*, 49(1):75–89, 2014.

[66] Brian Scassellati, Henny Admoni, and Maja Matarić. Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–294, 2012.

[67] Brian Scassellati, Laura Boccanfuso, Chien-Ming Huang, Marilena Mademtzi, Meiying Qin, Nicole Salomons, Pamela Ventola, and Frederick Shic. Improving social skills in children with asd using a long-term, in-home social robot. *Science Robotics*, 3(21), 2018.

[68] Walter Schneider and Arthur D Fisk. Attention theory and mechanisms for skilled performance. In *Advances in psychology*, volume 12, pages 119–143. Elsevier, 1983.

[69] Rebecca A Schwartz-Mette and Amanda J Rose. Conversational self-focus in adolescent friendships: Observational assessment of an interpersonal process and relations with internalizing symptoms and friendship quality. *Journal of Social and Clinical Psychology*, 28(10):1263–1297, 2009.

[70] Kate Silver and Sarah Parsons. Perspectives of autistic adults on the strategies that help or hinder successful conversations. *Autism & Developmental Language Impairments*, 7:23969415221101113, 2022.

[71] Susan H Spence. Social skills training with children and young people: Theory, evidence and practice. *Child and adolescent mental health*, 8(2):84–96, 2003.

[72] Sudha M Srinivasan, Isabel K Park, Linda B Neelly, and Anjana N Bhat. A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with autism spectrum disorder (asd). *Research in autism spectrum disorders*, 18:51–63, 2015.

[73] Stanford Artificial Intelligence Laboratory et al. Robotic operating system. URL https://www.ros.org.

[74] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[75] Tineke Backer van Ommeren, Sander Begeer, Anke M Scheeren, and Hans M Koot. Measuring reciprocity in high functioning children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42:1001–1010, 2012.

[76] Eleanor Watson, Thiago Viana, and Shujun Zhang. Machine learning driven developments in behavioral annotation: A recent historical review. *International Journal of Social Robotics*, pages 1–14, 2024.

[77] Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[78] Lauryn Wermer, Matthew E Brock, and Rachel L Seaman. Efficacy of a teacher training a paraprofessional to promote communication for a student with autism and complex communication needs. *Focus on Autism and Other Developmental Disabilities*, 33(4):217–226, 2018.

[79] Cheong Ying Sng, Mark Carter, and Jennifer Stephenson. A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism & Developmental Language Impairments*, 3:2396941518803806, 2018.

[80] Barry J Zimmerman. Development and adaptation of expertise: The role of self-regulatory processes and beliefs. *The Cambridge handbook of expertise and expert performance*, 186:705–722, 2006.