Chapter

Exploring Key Challenges in Child-Robot Interaction Using *Haru4Kids*: Engagement, Language Understanding, and Privacy

Gonzalo A. García, Leigh M. Levinson, Guillermo Pérez, Manuel Castro, José Gabriel Amores, Gloria Álvarez, Randy Gomez and Selma Šabanović

Abstract

This chapter examines three critical challenges in long-term child-robot interaction in the home, once the novelty effect has faded away: engagement, language understanding, and privacy concerns. The study used the *Haru4Kids* (H4K) platform, a child-oriented family robot simulator that features a rotating iPad-based interface offering interactive activities. This platform facilitated sustainable child-robot interactions in family environments, as it offered an interactive platform while also allowing us to assess user engagement and behavior throughout each interaction. Over two weeks, seven families in Southern Spain cohabitated with H4K. The study provides comprehensive insights into user engagement by integrating user logs, annotated images, and verbal interaction analysis. The image-based engagement was assessed via an innovative Engagement Level Metric, which we used to estimate which activities the children found most engaging. A natural language processing analysis revealed that mixed-initiative dialogs enhanced user agency over time, shifting interactions from system-driven to user-driven. Privacy concerns varied between children and parents, with children showing more hesitancy toward third-party data sharing. This chapter offers valuable design recommendations for future child-robot interaction platforms, emphasizing personalization, transparent data practices, and diverse activity offerings. For researchers and developers, it underscores the importance of addressing dynamic and multi-modal engagement and privacy concerns in realistic, unsupervised settings.

Keywords: child-robot interaction, engagement estimation, natural language processing, dialog management, privacy concerns, study *in the wild*, novelty effect

1. Introduction

Child-robot interaction (CRI) is a field within human-robot interaction that is growing in popularity and interest. Robots are increasingly being developed for children in influence areas like education [1, 2] and therapy [3]. However, much CRI research is limited to controlled settings due to the absence of robust, unsupervised platforms. Unlike controlled lab environments, “in-the-wild” environments, like a family home, allow children to interact with social robots voluntarily and in more dynamic ways [4]. This chapter introduces the *Haru4Kids* (H4K) platform, designed intentionally as a test-bed to evaluate children’s acceptance of robot *Haru* [5], particularly focusing on how children’s engagement with H4K changes across activities and over time, and on gauging families’ concerns about privacy.

Our work builds on previous studies involving Haru in child-robot interaction and aligns with UNICEF’s *Policy Guidance*,[1] which addresses children’s fundamental rights with AI and robots, including their right to privacy and fairness [6]. Designing a robotic system in line with a child’s rights requires not only careful consideration of the contexts in which the robot is to be used but also the guiding framework and scope of the interactions. It also demands careful consideration of the design features that will sustain long-term and satisfying CRI.

In order to achieve these goals, H4K features an avatar of the robot *Haru* [5, 7] displayed on an iPad held by a rotating stand that orients the iPad to face the child as they move. The design principle guiding Haru emphasizes a balance among human expectations, appearance, and functionality [5, 8]. This design principle takes a more holistic approach to CRI, favoring a variety of research methods and study designs to learn about it.

In this chapter, we integrate the findings from three studies: (1) takeaways from children’s engagement estimation by visual clues [9]; (2) different dialog strategies (user-driven, system-driven, and mixed-initiative) and evaluation of common built-in language intents of the system [10]; and (3) the pre- and post-perceptions of parents and children of information sharing and privacy concerns with robots in the home [11]. This work reflects on the fundamental preconditions for long-term child-robot interactions in dynamic environments like the home, including robot vision-based measurement of a child’s engagement with the system, improving child-robot dialogs, and aligning robotic use with familial comfort with sharing information. All of these components represented in our work are understudied in child-robot interaction (CRI), although they add great value as foundations for more sustainable integration of robots in children’s spaces.

2. Background

2.1 Long-term child-robot interaction

Cohabitating robots have been introduced to the home for a variety of purposes, such as entertainment, learning, healthcare, and companionship [12]. In the pursuit of measuring a *successful* and acceptable integration into children’s spaces, the field looks at how robots are used over longer periods. Defining “long-term” depends on interaction frequency, quality, capability, diversity, and scope [13, 14].

The kinds of benefits that sustain children’s interaction over time can include hedonic enjoyment or utilitarian benefits, like learning [15]. Additionally, the perception of a physical robot’s social presence and ability to exhibit expressive motion

---

[1] https://www.unicef.org/globalinsight/media/2356/file.

enhances user engagement compared to virtual agents [16, 17]. The personalized nature of robots is also a desirable feature and can enhance long-term engagement [18]. Furthermore, children are interested in robots that are expressive across verbal and non-verbal channels of communication, customizable, relatable, and approachable [19].

While establishing relationships enhances sustained interactions [20], doing so in the long term is challenging. Long-term studies show that many robotic platforms cannot overcome the *novelty effect*—the amplified but temporary excitement and use of a new device [12, 13]. As such, it is increasingly important for the field to define and measure different channels of successful interaction. In the following sections, we will expand on methods for measuring non-verbal and verbal forms of engagement with a robotic platform, and also the privacy concerns that may hamper even the short-term use and acceptance of robots.

2.2 Defining and measuring engagement

As already discussed, engaging and developmentally appropriate activities can support sustained interactions tailored to individual needs [14]. However, measuring the engagement of a user during an interaction is not a trivial task. Engagement, as defined by O'Brien [21], is the cognitive, affective, and behavioral investment in digital interaction, measurable through self-reports like the User Engagement Scale [22] and automatic estimation methods. Engagement with neurodivergent populations further suggests diversifying "engagement" across contexts, as it can support more personalized interactions over time [23]. The dynamic nature of engagement is also found in the home, for instance, with extended studies of home-based robots as reading companions demonstrating children's evolving engagement over several months [24, 25].

Currently, the most reliable engagement measurement methods rely on physiological signals [26], but these are intrusive and cumbersome for children. Other measurements of engagement are visual-based or multi-modal, which are unobtrusive and cost-effective [27], such as the recent one based on thermal imaging [28]. The existing automatic engagement estimation methods can be divided into those based on feature extraction and the ones that perform automatic prediction [29].

In our work, we look at visually-based engagement, including automatic face-angle tracking and hand-annotated engagement levels, which provide high resolution with minimal bias [30, 31]. Additionally, we explored engagement in CRI through dialog analysis, which is more objective, but less detailed.

Our approach uses a human-in-the-loop model, which integrates the feedback of human annotators for long-term engagement measurement in CRI. In this way, we take advantage of the natural human skill of *face reading* [32].

2.3 NLP as a key factor for engaging interactions

Dialog, enabled by Natural Language Processing (NLP), is a key way of achieving verbal engagement with young users. NLP has advanced significantly, making it possible for robotic systems to carry out an increasingly natural dialog with users. However, children's smaller vocal tracts and evolving language use create distinct difficulties in speech recognition [33, 34]. The lack of large, children-specific datasets further complicates speech recognition development, which is generally dominated by large corporations [35]. Children's imaginative language, ungrammatical phrases, and unique interests pose additional challenges [36]. Our work examines these issues in children's homes, an under-researched setting in CRI studies [37].

Unlike embodied robots, conversational agents have been more extensively studied in the home. Research on Alexa (Amazon), for example, explores breakdowns and repair strategies in communication [38] and analyzes language processing for joint reading [39]. These findings underscore the need for adaptive interaction models in child-robot dialog systems. As such, we pursue not only visual engagement analysis but also analyze the flow of dialogs with children and ways to improve verbal engagement with at-home robots.

2.4 Family privacy concerns

Though it is an understudied topic—surely due to its intrinsic difficulty to grasp and assess—ensuring privacy with at-home and child-centered robots is desired by family users [18]. Children's right to privacy builds from the general definition of privacy as the right to control what a user shares with another agent and how accessible what was shared is to others [40]. Dimensions of privacy also consider the control over being physically or psychologically alone [41]. As the kind of information that is shared varies in nature, it is expected that a user's willingness to share information or their privacy concerns will change based on the change of what is shared, whom it is being shared with, and how it is shared. This contextual nature of comfort is named by the framework of *contextual integrity* [42]. Within the context of social robots, we focus on informational, physical, social, and psychological dimensions of privacy given the heightened emotional and functional cues leveraged during interaction [43].

As a user's engagement shifts dynamically, their fluid relationship with robots across different contexts leads them to implicitly and explicitly navigate boundaries that they and their families establish. This boundary management has been described by the *communication privacy management theory*, which recognizes how users make others "co-owners" of their information depending on when they grant them access to information under their control [44]. In the case of families, pre-existing power structures, systematic hierarchies, and emotional relationships add a layer of complexity to privacy concerns [45].

These concerns can also vary widely across user demographics. Individual and generational differences affect how families establish, maintain, and amend these boundaries as they have done with other technologies in the past [46]. The privacy paradox with robots also recognizes how perceptions and expressed discomforts differ from actual user behavior [47]. All this context-driven variability in the management of privacy inspired our study of how parents and children conceptualize their privacy with a cohabitating social robot. Furthermore, studying familial privacy concerns will help researchers navigate the best contexts for the most comfortable engagement with these systems in the home.

3. Participating families and methods

3.1 Participant families

Seven families from Southern Spain participated in our study, recruited through convenience sampling. The study was approved by Indiana University's Institutional

Review Board #14363, with caregivers providing written consent and children (aged seven and older) giving written assent. In total, H4K engaged with 14 children (nine male) between ages 6 and 13 (mean age = 9.6). This sample size aligns with prior exploratory CRI studies, which often involve small samples, as noted in related works (e.g., [48–50]). While one family had an only child, the other families included at least two children. Accounting for periods when children were away, H4K cohabitated with each family for an average of 20 days ($\sigma = 5.9$, max = 27), resulting in a net period of two weeks per family. Researchers visited each home to install and set up H4K, explain privacy configurations to caregivers, and introduce the platform to family members. Children were then free to explore H4K independently.

3.2 *Haru4Kids* platform

H4K consists of a Haru avatar displayed on an iPad supported by a rotating stand (**Figure 1a**), using Apple libraries across seven modules: User Interface, Conversation Manager, Vision Manager, Central Controller, Settings Manager, Logging, and Stand Manager (**Figure 1b**). The Conversation Manager leverages cloud-based services for certain functions, while the Central Controller coordinates app behaviors and cloud access as needed. H4K's rotating stand augments its embodied presence, differentiating it from other virtual or voice-based systems.

H4K offers several activities designed for engagement and cognitive stimulation: *Storytelling*, *Gusano Loco* (a humorous, word-based storytelling activity), *Detectives* (word matching), *Would You Rather … ?* (choosing between silly options), and *Jokes*. These activities, created with neuropsychologists and speech therapists, were designed to stimulate different cognitive functions in children (for further details, see [9]).

3.3 Data collection

We collected data before, during, and after cohabitation. Images, audio, and logs were collected throughout the interaction. For a more holistic understanding of information sharing and general feedback on using the platform, we interviewed family

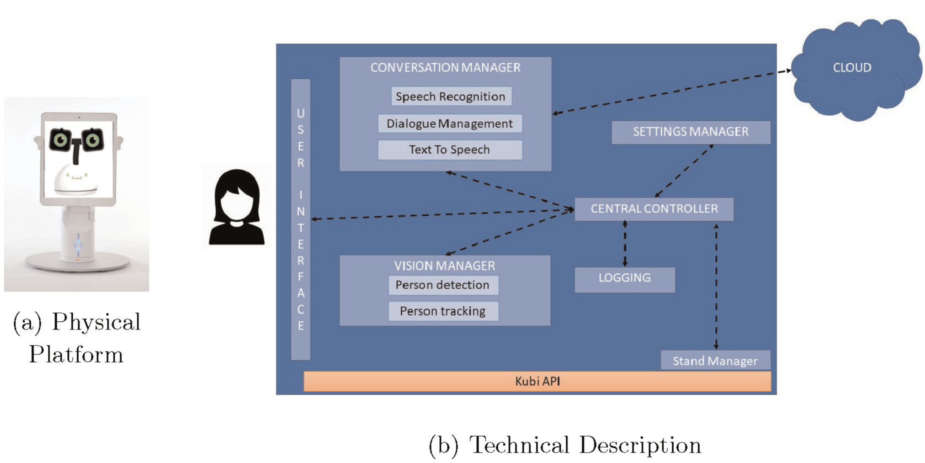

Figure 1.
Haru4Kids*: (a) iPad supported by rotating-tabletop stand, and (b) high-level technical description of the system.*

members before and after they interacted with H4K. Below, we describe the data collected in greater depth. All data was stored on the iPad locally and collected when the robots were retrieved from families' homes.

3.3.1 Experiment logistics

The families were recruited among acquaintances of one of the co-authors. The exact same protocol was followed with all the families: first, we made an appointment for deployment by e-mail. All deployments were carried out within the same week to keep the experimental conditions as similar as possible among the households. During the deployment, the same two researchers followed an identical process in all the homes: provide general explanations; provide explanations of the research consent, assent, and of the *General Data Protection Regulation* (GDPR) documents, first to parents, then to children: allow participants to read and sign the consent, assent, and GDPR documents; collect demographic data; complete pre-interview (parents, then children, separately); place the platform where the family indicates; set up the system; do an initial check; give brief instructions to family about how to use the system. Participants were given the contact information of the two researchers in charge of the experiment and encouraged to reach out if there were any issues.

After two net weeks, the researchers picked up the H4Ks and conducted a post-interview with the family members. The data was uploaded to the researchers' local server, which is highly protected and is only accessible to the researchers. The data were then permanently and irrecoverably deleted from the iPads.

3.3.2 User logs

All application events were logged both on the iPad and—anonymously—in the cloud (AWS Cloudwatch). Log entries included timestamps, event types (e.g., actions by H4K or user requests), user head angles (from Apple's Vision framework), and rotating stand movements, as these can correlate with engagement levels [51]. These logs provide a continuous record of user-robot interaction, supporting subsequent analysis of dialog quality [52].

3.3.3 Image capture

To gauge user engagement, we captured images of users' faces at one frame per second during interactions from six of the seven families who consented to image collection. Images were displayed in real time on the iPad for transparency.

3.3.4 Audio and conversation manager

H4K's Conversation Manager, in charge of handling the components needed for voice-based interactions, integrates Apple libraries for speech recognition and its synthesis (Text to Speech). The Dialog Management component is a custom-built container hosted on AWS that allows complex, mixed-initiative dialogs beyond standard chatbot interactions and features error-handling capabilities like *Repeat*, *Help*, *Background*, and *Sleep*, providing a robust interaction experience even in error-prone, unsupervised home environments.

Conversation Manager is also in charge of the audio signal recording, which was limited to interactions where children responded verbally, with data stored locally and never uploaded to the cloud. Families retained full control over data deletion at any stage of the experiment (even after it was finished), respecting privacy guidelines.

3.3.5 Pre- and post-interview data

To better understand caregiver and child experiences and concerns with a cohabitating robot, we also utilized pre- and post-interviews. In a 30-minute pre-interview, researchers asked children about their first impressions and expectations for interactions with the robot. Notably, we asked multiple-item questions about information sharing with a social robot (e.g., *"What kind of information would you feel comfortable showing or telling with a robot?"*) and about sharing information with different third parties through the robot (e.g., *"What information would you feel comfortable having the robot share with your teacher/friend/sibling/parent/robot creator?"*). Parents were asked about their own comfort with sharing information with the robot and comfort with their child interacting and sharing information with the robot and third parties (total of 7 closed, information-sharing questions with 12 items each).

At the end of the trial period, a 30–45 minute post-interview was conducted with all family members, who were asked the same multi-item questions for a pre-post analysis. Kids were also asked about what they enjoyed, did not enjoy, or would like Haru to do in the future. Parents were also asked about their impressions of cohabitation and feedback for the roboticists. Furthermore, we asked each participating child to draw Haru, so that we could gauge what they perceived as the most salient features of the robot.

For more specific details about the methods here presented, please refer to our publication on this study [11].

3.4 Data analysis

3.4.1 Interview analysis

Comfort levels and other quantitative measures were analyzed statistically using t-tests and a p-value of 0.05. Qualitative answers are described generally and reported without thorough thematic analysis. These interviews were mainly used for internal feedback and to guide the engagement analyses here described.

3.4.2 Measures for engagement estimation based on usage and dialog

From the log files, we want to call attention to the extraction of the usage profile of each family and user reactions to the robot. Specifically, for each family and each day of the experiment, we calculated the number of sessions per day and their durations based on their starting and ending timestamps. Additionally, we determined the number of times each activity was executed. We also analyzed whether it was actively requested by the user (user-driven) or proposed by the robot (system-driven) and whether it was completed or aborted by the user. From the log entries related to dialogs, we analyzed the users' reactions, or lack thereof, to the robot's participation cues to the user, so that we obtained an additional estimation of their level of attention and engagement.

3.4.3 Engagement profile estimation from users' pictures analysis

An *Engagement Level Metric* (ELM) was developed based on hand-annotated user photos categorized into four levels of engagement. Annotators, trained in a seminar and equipped with a guide, assessed a total of around 20 K collected images using the ELM, which follows descriptions from prior studies [53, 54]. Annotators were provided only with facial images, ensuring unbiased assessments. Our annotation support application monitored annotator attentiveness by inserting test images and tracking response times.

3.4.4 Additional measures for engagement estimation

- *User head angle (UHA)*: Correlation studies were conducted between UHA components and the ELM, as previous literature has associated engagement with angles within $\pm 12^{\circ}$, indicating user focus on the screen [51].
- *Stand cumulative angle (SCA)*: This metric measures the stand's movement, assuming a higher SCA indicates decreased engagement.

3.4.5 Assessing annotation quality and reliability analysis

Annotation reliability was assessed using inter-rater coefficients such as Krippendorff's α, Cohen's κ, and Cronbach's α [55–57], as well as general agreement measures like Pearson's correlation and Root Mean Square Deviation (RMSD), which was normalized by $(y_{max} - y_{min})$, 3 in our case.

4. Results

4.1 General use over time

Figure 2 shows the cumulative usage time across families over two weeks. Individual families' usage is shown in the inset of **Figure 2**. Overall usage was low, yielding a total of 18.37 hours, and quite variable both in usage length, with one family achieving 15 days of net use and another just three; and in total usage time, with a maximum of 4.92 hours and a minimum of 53 minutes. After day #10, only one family continued interacting with the robot, showing how once the novelty effect fades away, the usage drops steeply.

Figure 3a illustrates usage by activity, showing how children preferred *Jokes*, *Storytelling*, and *Gusano Loco*, which is aligned with self-reported preferences [11]. The interviews revealed results in line with this feedback. Of all the activities, children expressed the most excitement over jokes and the least excitement for Haru's storytelling.

In general, most interactions were user-driven (red bars), with children increasing activity requests from 33% on day one to 85% by the last day as shown in **Figure 3b**. This result shows the importance of the mixed-initiative approach, which allows the children to either passively wait for the robot's proposals, or to actively request a specific activity. Further details about this approach will be discussed in Section 4.4.2.

4.2 Visually measured engagement profile

Children's attentiveness was analyzed using verbal responses to H4K's prompts, with results shown in Section 4.4. Engagement levels were primarily assessed through

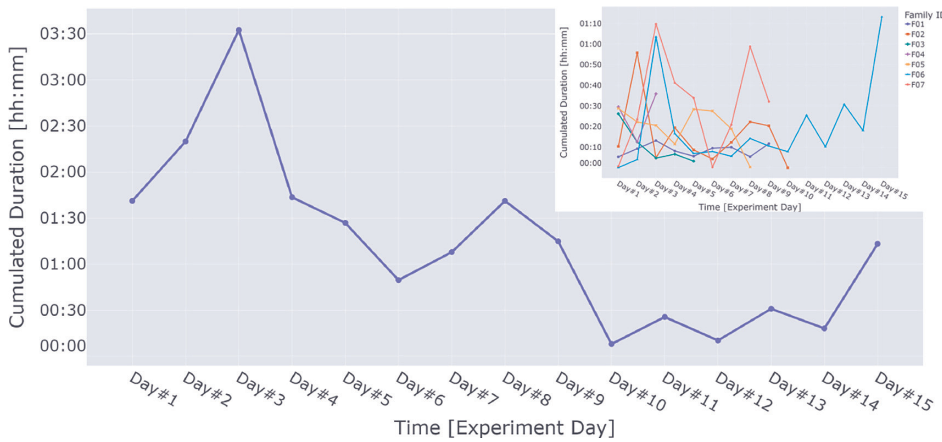

Figure 2.
Evolution of the average, global time usage by day of experiment. Inset: individual family behavior.

Figure 3.
Comparison of activity usage and user-initiated activities over time. (a) Comparison of activities usage. (b) User-initiated activities percentage over time.

hand-labeled interaction photos, though these were less objective than dialog-based metrics.

After annotating, the inter-annotator agreement was assessed, resulting in the exclusion of two annotators with low agreement scores. Inter-rater agreement level was mixed: Krippendorff's α indicated low reliability, while Cohen's κ suggested fair-to-moderate agreement, and Cronbach's α indicated good agreement. Those measures are known to underestimate agreement when one label dominates [58]. On the other hand, Pearson's r scores (0.43–0.55) were strong by psychological research standards [59]. Collectively, these values support the ELM's validity as an engagement estimate, given also the high RMSD-based metric (71–74%).

Based on those fair-to-moderate agreement values among the remaining 10 annotators, the data provide reasonable confidence in the labels used for engagement analysis. To further ensure the validity of the manual labeling, we selected from the labels for each set only those in which at least three annotators agreed in its ELM, which was 79% of the pictures. Afterward, we considered the ELM of each picture the most-voted label, $\overline{ELM}_3$.

Those metrics show that activities like *Detective*, *Would You Rather*, and *Gusano Loco* had slightly higher engagement. Engagement declined over time, but the differences among activities remained statistically significant, with *Detective* ranking highest and *Story* lowest, which is in line with the results obtained from user logs and the interviews.

4.3 Other engagement metrics: SCA and UHA

SCA correlations were minimal (0.07 for yaw, -0.03 for pitch), while UHA absolute yaw angle (head rotation by the neck axis, eye gaze moving toward a shoulder) showed a notable -0.43 correlation with $\overline{ELM}_3$. These results indicate that perceived engagement decreased as children looked away from the screen.

Overall, we describe how the annotated pictures amidst other visually-based indicators provide a useful and trustworthy estimation of children's engagement. In the following section, we will analyze child-user dialogs with measures that do not involve the intervention of human annotators. While the automated process makes them more objective, it may miss some of the nuances a human can extract from a face image. We recognize these differences and present the general takeaways from the study.

4.4 Child-robot dialogs and verbal engagement

4.4.1 Children's language understanding in the wild

The Automatic Speech Recognition (ASR) outputs were transcribed and analyzed by humans, yielding an overall Word Error Rate (WER) of 0.077, with multiple-choice questions performing best (WER = 0.07) and open-ended questions worse (WER = 0.11). Hence, ASR showed robust performance, even though some of the participating children have a strong Southern accent that makes their speech more difficult to interpret, as it differs from the *standard* pronunciation. In any case, the accuracy of the ASR could be further improved by optimizing for common names or fine-tuning the models.

Table 1 details intent recognition metrics, with high performance across all categories (F ($\beta = 0.5$) $\geq$ 0.97). However, low occurrence counts for some intents (e.g., *Repeat*) may affect generalizability. The results exclude recognition errors, silences, and out-of-scope (OoS) responses, but these collectively accounted for 24% of responses in yes/no and multiple-choice questions.

Handling OoS inputs effectively is vital for user experience [60], with Dialogflow's fallback intent yielding a performance of $Precision = 0.90$ and $Recall = 0.98$.

Error rates declined slightly over time, with "silence" rates decreasing markedly toward the end of the trial, suggesting children became more familiar with how to respond to the robot's prompts.

4.4.2 Children-robot dialog management in the wild

Standard dialog metrics, such as average user turns (5.25) and session duration (1.92 minutes), varied between families, with one family averaging almost five

	yes/no	multiple	sleep	repeat	activity
Precision	0.992	0.985	1.000	1.000	0.993
Recall	0.985	0.958	0.886	1.000	0.987
F ($\beta = 1$)	0.989	0.971	0.939	1.00	0.990
F ($\beta = 0.5$)	0.991	0.980	0.975	1.00	0.992

Table 1.
Intent recognition performance.

minutes per session. As already shown in **Figure 3b**, user-requested activities increased over time, indicating growing familiarity with the system.

Domain-independent intents (e.g., *Repeat* and *Help*) were infrequently used, except for *Background*, which was frequently triggered (1.31 times per session). Most sessions ended automatically when users disengaged, with fewer instances of using the *Sleep* intent manually. The NLU layer performed well in general, with pre-defined intents achieving F-scores of 0.98 or higher. While recognition issues affected 10% of inputs, they did not decrease over time, suggesting this may be a stable limit for our current setup.

Within this work, our takeaways are two-fold: (1) there is a clear benefit of mixed-initiative dialog, allowing users to shift from predominantly system-driven interactions to increasingly user-driven interactions by the experiment's end; and (2) the observation that functionalities like *Help* and *Repeat* require explicit user training, while a fallback *Background* strategy is critical in maintaining flow in conversational applications.

Further, H4K offers valuable dialog statistics on turn-taking, error handling, and dialog closure (see details in [10]) that could inform future CRI designs.

4.5 Dynamic comfort in sharing information in the home

To complement the above metrics of visual and verbal engagement with the robot, we report the evolution of the family members' general comfort of having Haru in the home. See **Table 2** which describes children's change in comfort in sharing

Info Category	General	Third Parties				
		Teacher	Friends	Siblings/ Cousins	Parents	Robot Creators
Third Party		74%	80%	83%	86%	40%
School Grades	57% $\downarrow_3$	100%	50%	50%	50%	21% $\downarrow_3$
Hobbies	100%	79% $\downarrow_1$	100%	93% $\downarrow_1$	86% $\downarrow_1$	86%
Conversations with Others	14% $\downarrow_2$	7% $\uparrow_1$	14% $\downarrow_1\uparrow_1$	14% $\downarrow_3\uparrow_1$	29% $\downarrow_3\uparrow_1$	14% $\downarrow_1$
Name	100%	100%	100%	100%	100%	42% $\downarrow_3$
Birthday	86% $\downarrow_2$	93% $\downarrow_1$	100%	100%	100%	21% $\downarrow_5$
Pets	100%	93%	100%	100%	100%	71% $\downarrow_1$
Family Info	43% $\downarrow_1$	35% $\downarrow_1$	50% $\downarrow_1$	100%	93% $\downarrow_1$	14% $\downarrow_2\uparrow_1$
Friend's Info	64% $\downarrow_3$	93% $\downarrow_1$	100%	50% $\downarrow_3$	79% $\downarrow_2$	42% $\downarrow_4$
Location	43% $\downarrow_3$	36% $\downarrow_1$	76% $\downarrow_1$	100%	100%	21%
Voice Recognition	100%	100%	100%	100%	100%	71% $\downarrow_4$
Face Recognition	93%	100%	100%	100%	100%	42% $\downarrow_3$
Images of family	57% $\downarrow_2$	42% $\downarrow_3$	57% $\downarrow_3\uparrow_1$	83% $\downarrow_1$	93% $\downarrow_1$	21% $\downarrow_2$

$\downarrow$ = *number of children who changed to discomfort in the post-interview.*
$\uparrow$ = *number of children who changed to comfort in the post-interview.*
Blue = 100% Comfort, Green = $\geq$50% comfortable, Red = <50% comfortable.

Table 2.
Child comfort with sharing information originally reported by Levinson et al., 2022 [11].

information. All children, regardless of previous exposure to a prototype, had high initial expectations for the robot's capabilities. As such, children reported disappointment when H4K struggled to understand their input, which was related to our NLP findings and impacted interaction frequency. There was also an overall decrease in comfort with sharing information, in line with this reported disappointment in the robot's capabilities.

Regarding privacy, only 14% of children felt comfortable sharing conversations with others with the robot, and 42% felt comfortable sharing location or family information with the robot. The most comfortable information shared included names, pets, and voices.

Post-interview results showed that 98% of adults were comfortable with children sharing data directly with them but were less at ease with sharing data with teachers (69%) and robot creators (67%), underscoring the contextual comfort with whom information is shared. In comparing children's pre- and post-interview results, a paired t-test determined that there was a significant difference before and after the cohabitation period ($t = 5.303$, $p < 0.001$). In comparison, a paired t-test of the adult's pre- and post-interview responses about information sharing was not significant ($t = 1.387$, $p = 0.22$). Therefore, children were more likely than parents to change their answers post-interaction than their caregivers. While this does not directly reflect on the privacy paradox, which identifies how user behavior differs from user preferences, it elucidates that younger children have more dynamic boundaries, justifying a greater need to be transparent with information sharing so they can make informed decisions on their robot use.

5. Discussion

To our knowledge, our work is the first one to provide such a multi-faceted study in CRI in the home environment. In the following discussion, we will situate our work within the research on CRI and the deployment of engaging systems.

5.1 Children's dialogs with robots

As Ljunglöf et al. [61] noted, speech recognition has historically been dialog systems' weak point, but our results align with recent improvements in ASR technology [62]. Although not yet at human-level accuracy, these results show notable progress, particularly for non-English languages like Spanish in real-world contexts.

Our reliable performance of speech recognition in this challenging scenario, similar to the accuracy rates reported in the study of Xu and Warschauer [39], indicates progress but leaves more to be desired. There are challenges unique to voice interfaces with children, such as children's reliance on and use of non-verbal gestures in communication [39], which motivates an enhanced focus on dialog management for child-robot interactions with multi-modal capabilities.

As for language understanding, although our NLU layer performed well in general, new NLUs based on current Large Language Models (LLMs) could overcome some limitations of more "classic" approaches like ours, as LLMs offer a deeper apprehension of user inputs. However, LLMs arise some issues such as privacy concerns and the possibility of *hallucinations* and non-proper answers.

Mixed-initiative dialogs allowed children to initially follow system-driven interactions and later take control to request specific activities, as shown in **Figure 3b**. This is a novelty with respect to existing systems, which are either fully passive—just answering the questions addressed to them by the user (e.g., *Alexa*)—or fully autonomous—guiding the user through the (maybe pre-established) dialog (e.g., a voice menu over a phone call). Mixed-initiative dialogs allow users to become more familiar with the system by the scaffolding of the guided dialog until they feel confident enough to lead the dialog.

5.2 A multi-modal engagement study

Our overall work is comprehensive in that it can take into account engagement measures across user logs and annotated images, dialog responses, and self-report interviews. Visual engagement estimation and interviews were performed on an individual's interaction data, whereas estimations by dialog and usage were rather measured family-wide. Our results found that generally, children and parents, while initially comfortable with sharing images during interactions, became slightly less comfortable after the interaction. However, they were more comfortable with voice recognition and the recording of audio data. While images may offer a rich and well-established source of information about the interaction, these results inspire future work on triangulating engagement across these modalities to best align data collection with familial comfort.

It is worth noting that the ELM ranking of activities $(\text{Detective} \gtrsim \text{WYR} \approx \text{Jokes} \gtrsim \textit{Gusano Loco} \gtrsim \text{Story})$ collected from images did not correspond directly to usage frequency from the user logs or self-reported enjoyment $(\text{Jokes} >> \text{Story} >> \textit{Gusano Loco} > \text{Detective} \approx \text{WYR})$. In this way, we identify an inequality between the activities most enjoyed as reported by children and their families, most frequently used, and those in which children appeared more engaged. Therefore, ELM could guide design for highly engaging activities, while usage profiles could identify the characteristics of popular, frequently used content. This approach offers a dual perspective for activity design in future CRI platforms.

This multi-modal approach also validated diverse kinds of engagement in CRI. The higher engagement levels of *Detective* could inform the design of future activities that emphasize problem-solving and interaction diversity. In fact, studies have highlighted that game-based problem-solving activities foster intrinsic motivation and sustained engagement in children during human-computer interaction scenarios, aligning with our findings [63]. Additionally, the decline in engagement after the novelty effect wanes indicates that personalization and adaptability—features where mixed-initiative dialogs play a role—are essential to sustaining interest. On the side of personalization, we highlight how our Engagement Level Metric (ELM) provided a robust framework for estimating engagement based on image annotations. While tools like ELAN [64] offer annotation capabilities, our custom tool was specifically designed for this study's requirements, including features for annotator performance monitoring in real time and future usability. In this way, the engagement metric is more personalized to the participating children.

Our results are also in line with other work that emphasizes the role of mixed-initiative for engaging human-robot interactions [65], particularly in the case of maintaining engagement over longer periods of time [66].

Furthermore, we realized that children are not always more engaged by the activities they requested the most, as at some ages, they like serial repetitions of known content [67]. Thus, we recognize that both the nature of the interaction and the content are important contributing factors toward engaging CRI.

5.3 Privacy and information sharing

Beyond takeaways surrounding the topic of engagement, we briefly report some of the privacy-related findings that establish a foundation for trusting and successful child-robot relationships. Survey responses revealed an overall decrease in comfort in sharing information with a robot, with children exhibiting more caution than parents around third-party data sharing. The dynamic shift and decrease after cohabitation also make space for the re-navigation of privacy boundaries in the home, in line with the privacy communication management theory proposed before [44]. In this, we also find that children are more aware of the privacy risks with technology than parents or other adults realize [40].

In particular, we found that voice recognition was the most accepted information category among all dimensions, especially sharing with robot creators. With this in mind, it becomes critical to shift the CRI research focus toward improving dialog-based NLP and using voice-based metrics to assess the quality of child-robot interactions.

5.4 Limitations and future work

Our findings align with past research indicating challenges in sustaining children's attention over time, as the novelty effect wears off [68]. Activity decreased gradually. Limited content and the lack of embodiment might have influenced engagement, with our app-based platform lacking the physical presence that some children might prefer in social agents [11]. While embodied agents often offer benefits over screen-based ones [69, 70], we used an avatar to allow faster, more flexible deployment in multiple homes. This setup successfully ensured technical consistency throughout the experiment, reducing the risk of hardware issues and enabling easy configuration adjustments without compromising activity design. Though our platform has a rotating stand providing a degree of embodiment and movement, we recognize that this may have been a factor in decreased interaction. In the future, we hope to best address this with more robust platforms that can be both consistent and embodied.

This study was limited by sample size, language, and geographic scope. A planned follow-up will expand these areas, including longer cohabitation, broader demographics, and more diverse activities. Key enhancements include adapting activities to various developmental stages and cultural contexts and introducing collaborative storytelling. We plan to develop additional educational resources and explore cohabitation with other populations, such as hospitalized children, to leverage H4K's cognitive benefits for diverse users [71].

The two-week cohabitation demonstrated initial user interest that tapered during the second week, likely due to the onset of summer break. Future deployments should address expectations with transparency and age-appropriate content. In addition, the nearly 20 K pictures annotated in this study could be used as a dataset for training artificial intelligent models aimed at automatic engagement estimation, even in real time, closing the loop and making it possible to adapt the behavior of the robot to the attention level detected in the child user.

The intents *Help* and *Repeat* were underused, likely due to limited awareness of their existence. Therefore, we will need to introduce these features more explicitly when deploying the robots in order to improve future interaction.

As for privacy issues, future research and developments will have to take into account new legislation such as the European *AI Act*,[2] which places limitations on the collection of some types of information, as well as the use of large language models (e.g., *Meta*'s *Llama* 3.2[3]). One possible workaround for this could be to create more complex and useful logs, in which interpretations of the acts performed by the user—rather than the recording of the acts themselves—will be locally recorded in the robot and then used by the researchers, who will thus not have access to identifiable, private information (audio or video signals). Such an approach is already in practice in devices like the one used by the USA national nonprofit *LENA* (*Language ENvironment Analysis*): a small wearable device (a "talk pedometer") that records a log of events (e.g., the number of turns in a dialog) rather than the dialogs themselves and that it is used for speech analysis of toddlers and children.

6. Conclusions

The Haru4Kids (H4K) platform presented in this chapter represents a significant contribution to the field of child-robot interaction (CRI), providing a robust and versatile tool for studying engagement, language understanding, and privacy concerns in real-world, uncontrolled settings. By integrating multi-modal engagement metrics—ranging from visual data and dialog analysis to self-reported feedback—this research underscores the complexity of fostering meaningful and sustained interactions between children and social robots.

Our findings reveal critical insights: children exhibit enthusiasm for cohabitating robots, though their engagement fluctuates based on activity type, system familiarity, and the waning novelty effect. Privacy concerns emerged as another pivotal dimension, with children and parents navigating shifting comfort levels over time. In our opinion, these observations highlight the need for adaptive and transparent CRI systems that address both ethical and technical challenges.

The study's novelty lies in its comprehensive approach: combining user-driven and system-driven dialogs through mixed-initiative strategies, leveraging annotated image-based engagement metrics, and exploring privacy dynamics in unsupervised family environments. These contributions provide a foundation for future advancements in CRI, particularly in personalizing interactions and ensuring ethical data practices. In addition, our results may be useful for studies in developmental psychology and user engagement, supporting a growing field of research that combines objective and subjective engagement measures to foster meaningful, interactive child-robot experiences.

The broader implications of this research lie in its potential to guide the development of incoming child-oriented systems that balance entertainment, education, and ethical considerations [72].

Looking ahead, we propose enhancing activity diversity and platform embodiment while expanding demographic and cultural diversity in future trials. We also emphasize the integration of advanced natural language processing systems to deepen interaction quality. This research not only advances scientific understanding of CRI but

---

[2] https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[3] https://www.llama.com/llama3_2/use-policy/.

also provides actionable insights for developing engaging, ethical, and inclusive child-robot interaction platforms for diverse real-world applications.

Acknowledgements

We would like to thank all the participating families, Sol Benitez Casaccia for helping with the translations throughout the process, and Sunzia Reza, Margaret Nigh, and Kaleb Lewellen for their help with content generation. The authors also thank Marta López, Ricardo Durán, Iván Delgado, José Andrés Millán, and Douwe Gelling for their technical essential contributions.

This publication is part of the project *PLEC2023-010251*, funded by MICIU/AEI/10.13039/501100011033.

This work was supported by *Honda Research Institute Japan* and by the Spanish Ministry of Science, Innovation and Universities under grants *PTQ2021-011726* and *PLEC2023-010251* funded by MICIU/AEI/10.13039/501100011033.

ChatGPT 4o (*OpenAI*) was used in the elaboration of this manuscript for style polishing, improving the tables, and summarization of some sections.

Author details

Gonzalo A. García[1,2]*, Leigh M. Levinson[3], Guillermo Pérez[1], Manuel Castro[1], José Gabriel Amores[4], Gloria Álvarez[4], Randy Gomez[5] and Selma Šabanović[3]

1 4i Intelligent Insights, Seville, Spain

2 UNIR - Universidad Internacional de La Rioja, Logroño, Spain

3 Indiana University Bloomington, Bloomington, USA

4 Universidad de Sevilla, Seville, Spain

5 Honda Research Institute, Wako, Japan

*Address all correspondence to: g.garcia@4i.ai

References

[1] Toh LP, Causo A, Tzuo PW, Chen IM, Yeo SH. A review on the use of robots in education and young children. Journal of Educational Technology and Society. 2016;**19**(2):148-163

[2] Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F. Social robots for education: A review. Science Robotics American Association for the Advancement of Science. 2018;**21**(3):eaat5954

[3] Cabibihan JJ, Javed H, Ang M, Aljunied SM. Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. International Journal of Social Robotics. 2013;**5**(4):593-618

[4] de Graaf M, Ben Allouch S, van Dijk J. Why Do they Refuse to Use my Robot? Reasons for Non-use Derived from a Long-Term Home Study. New York, NY: ACM Press; 2017. pp. 224-233

[5] Gomez R, Szapiro D, Galindo K, Nakamura K. Haru: Hardware design of an experimental tabletop robot assistant. International Conference on Human-Robot Interaction (HRI). 2018:233-240

[6] Charisi V, Imai T, Rinta T, Nakhayenze JM, Gomez R. Exploring the concept of fairness in everyday, imaginary and robot scenarios: A cross-cultural study with children in Japan and Uganda. Interaction Design and Children. 2021;**24**:532-536

[7] Sandry E, Gomez R, Nakamura K. Art, design and communication theory in creating the communicative social robot ‘Haru’. Frontiers in Robotics and AI. 2021;**8**:577107

[8] Ackerman E. Haru: An experimental social robot from honda research. IEEE Spectrum. 2018. Available from: https://spectrum.ieee.org/haru-an-experimental-social-robot-from-honda-research

[9] García GA, Pérez G, Levinson L, Amores JG, Alvarez-Benito G, Castro-Malet M, et al. Living with Haru4Kids: Study on children’s activity and engagement in a family-robot cohabitation scenario. In: 32nd IEEE Conference on Robot and Human Interactive Communication (RO-MAN). New York, USA; IEEE; 2023. pp. 1428-1435

[10] Pérez G, García GA, Castro M, Castaño M, De Quevedo MJLG, Durán R, et al. Child-robot conversation in the wild wild home: A language processing user study. In: 2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). New York, USA; IEEE; 2023. pp. 1436-1442

[11] Levinson L, García GA, Pérez G, Alvarez-Benito G, Amores JG, Castaño-Ocaña M, et al. Living with Haru4Kids: Child and parent perceptions of a co-habitation robot for children. In: Social Robotics: 14th International Conference, ICSR 2022, Florence, Italy, December 13–16, 2022, Proceedings, Part II. Cham: Springer Nature Switzerland; 2023. pp. 54-63. DOI: 10.1007/978-3-031-24670-8_6

[12] Leite I, Martinho C, Paiva A. Social robots for long-term interaction: A survey. International Journal of Social Robotics. 2013;**5**(2):291-308

[13] Ahmad MI, Mubin O, Orlando J. Adaptive social robot for sustaining social engagement during long-term children–robot interaction. International Journal of Human Computer Interaction. 2017;**33**(12):943-962

[14] Cagiltay B, White N, Ibtasar R, Mutlu B, Michaelis J. Understanding factors that shape children's long-term engagement with an in-home learning companion robot. In: Proceedings of the 21st Annual ACM Interaction Design and Children Conference. New York, USA: ACM; 2022. pp. 362-373

[15] De Jong C, Peter J, Kühne R, Van Straten C, Barco A. Exploring children's beliefs for adoption or rejection of domestic social robots. In: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). New York, USA; IEEE; 2021. pp. 598-603

[16] Leite I, Martinho C, Pereira A, Paiva A. As time goes by: Long-term evaluation of social presence in robotic companions. In: RO-MAN 2009 - the 18th IEEE International Symposium on Robot and Human Interactive Communication. Toyama, Japan: IEEE; 2009. pp. 669-674

[17] Knight H, Simmons R. Expressive motion with x, y and theta: Laban effort features for mobile robots. In: 23rd IEEE International Symposium on Robot and Human Interactive Communication. New York, USA; IEEE; 2014. pp. 267-273

[18] Cagiltay B, Ho HR, Michaelis JE, Mutlu B. Investigating family perceptions and design preferences for an in-home robot. In: Proceedings of the Interaction Design and Children Conference. New York, USA: Association for Computing Machinery (ACM); 2020. pp. 229-242

[19] Levinson L, Gomez R, Šabanović S. Bow ties & colorful eyes: Centering youth designs of social robots. In: 2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN). IEEE; 2024. pp. 1667-1674

[20] Kanda T, Sato R, Saiwaki N, Ishiguro H. A two-month field trial in an elementary school for long-term human–robot interaction. IEEE Transactions on Robotics. 2007;**23**(5):962-971

[21] O'Brien H. Theoretical perspectives on user engagement. In: O'Brien H, Cairns P, editors. Why Engagement Matters. Cham: Springer; 2016

[22] O'Brien HL, Cairns P, Hall M. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. International Journal of Human-Computer Studies. 2018;**112**: 28-39

[23] Papakostas GA, Sidiropoulos GK, Lytridis C, Bazinas C, Kaburlasos VG, Kourampa E, et al. Estimating children engagement interacting with robots in special education using machine learning. Mathematical Problems in Engineering. 2021;**1**:9955212

[24] Zhao Z, McEwen R. "Let's read a book together" a long-term study on the usage of pre-school children with their home companion robot. In: Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction. New York, USA: Association for Computing Machinery (ACM); 2022. pp. 24-32

[25] Michaelis JE, Mutlu B. Reading socially: Transforming the in-home reading experience with a learning-companion robot. Science robotics. 2018; **3**(21):eaat5999

[26] Alimardani M, van den Braak S, Jouen AL, Matsunaka R, Hiraki K. Assessment of engagement and learning during child-robot interaction using EEG signals. In: Social Robotics: 13th International Conference, ICSR 2021, Singapore, Singapore, November 10–13,

2021, Proceedings 13. New York, USA: Springer International Publishing; 2021. pp. 671-682

[27] Dewan M, Murshed M, Lin F. Engagement detection in online learning: A review. Smart Learning Environments. 2019;**6**(1):1-20

[28] Levinson L, Kaufman Z, Karim A, Huang A, Gomez R, Sabanovic S. The nose knows: Using thermal imaging to approximate children's engagement with robots. In: Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. New York, USA: Association for Computing Machinery (ACM); 2024. pp. 669-673

[29] Sorrentino A, Fiorini L, Cavallo F. From the definition to the automatic assessment of engagement in human–robot interaction: A systematic review. International Journal of Social Robotics. 2024;**16**:1641-1663

[30] Pabba C, Kumar P. An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. Expert Systems. 2022;**39**(1):e12839

[31] Charisi V, Davison D, Reidsma D, Evers V. Evaluation methods for user-centered child-robot interaction. In: 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). New York, USA; IEEE; 2016. pp. 545-550

[32] Wierzbicka A. Reading human faces: Emotion components and universal semantics. Pragmatics & Cognition. 1993;**1**(1):1-23

[33] Potamianos A, Narayanan S, Lee S. Automatic speech recognition for children. In: Fifth European Conference on Speech Communication and Technology. Grenoble, France: International Speech Communication Association (ISCA); 1997

[34] Russell M, D'Arcy S. Challenges for Computer Recognition of children's Speech. Workshop on Speech and Language Technology in Education. Grenoble, France: International Speech Communication Association (ISCA); 2007

[35] Liao H, Pundak G, Siohan O, Carroll M, Coccaro N, Jiang QM et al. Large Vocabulary Automatic Speech Recognition for Children. New York, USA; IEEE; 2015

[36] Gray SS, Willett D, Lu J, Pinto J, Maergner P, Bodenstab N. Child automatic speech recognition for US English: Child interaction with living-room-electronic-devices. WOCCI. 2014:21-26

[37] Van Straten CL, Peter J, Kühne R. Child-robot relationship formation: A narrative review of empirical research. International Journal of Social Robotics. 2020;**12**(2):325-344

[38] Beneteau E, Richards OK, Zhang M, Kientz JA, Yip J, Hiniker A. Communication breakdowns between families and Alexa. In: Proceedings of CHI Conference on Human Factors in Computing Systems. New York, USA: Association for Computing Machinery (ACM); 2019. pp. 1-13

[39] Xu Y, Warschauer M. Exploring young children's engagement in joint reading with a conversational agent. In: Proceedings of the Interaction Design and Children Conference. New York, USA: Association for Computing Machinery (ACM); 2020. pp. 216-228

[40] Zhao J, Wang G, Dally C, Slovak P, Edbrooke-Childs J, Van Kleek M, et al. I make up a silly name' understanding Children's perception of privacy risks

online. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. New York, USA: Association for Computing Machinery (ACM); 2019. pp. 1-13

[41] Gavison R. Privacy and the limits of law. The Yale Law Journal. 1980;**89**(3): 421-471

[42] Nissenbaum H. Privacy as contextual integrity. Washington Law Review. 2004;**79**:119

[43] Lutz C, Schöttler M, Hoffmann CP. The privacy implications of social robots: Scoping review and expert interviews. Mobile Media & Communication. 2019;**7**(3):412-434

[44] Petronio S. Communication privacy management theory: Understanding families. In: Engaging Theories in Family Communication. Abingdon, UK: Routledge; 2017. pp. 87-97

[45] Petronio S, Child JT. Conceptualization and operationalization: Utility of communication privacy management theory. Current Opinion in Psychology. 2020;**31**:76-82

[46] Miltgen CL, Peyrat-Guillard D. Cultural and generational influences on privacy concerns: A qualitative study in seven European countries. European Journal of Information Systems. 2014; **23**(2):103-125

[47] Lutz C, Tamó-Larrieux A. The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. Human-Machine Communication. 2020;**1**:87-111

[48] Charisi V, Gomez E, Mier G, Merino L, Gomez R. Child-robot collaborative problem-solving and the importance of child's voluntary interaction: A developmental perspective. Front Robot AI. 2020;**7**:15

[49] Leyzberg D, Ramachandran A, Scassellati B. The effect of personalization in longer-term robot tutoring. ACM Transactions on Human-Robot Interaction (THRI). 2018;**7**(3): 1-19

[50] Kory-Westlund JM, Breazeal C. A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. Frontiers in Robotics and AI. 2019;**6**:81

[51] Pan Y, Hirokawa M, Suzuki K. Measuring K-degree facial interaction between robot and children with autism spectrum disorders. In: 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). New York, USA; IEEE; 2015. pp. 48-53

[52] Esposito R, Fracasso F, Limosani R, Onofrio GD, Sancarlo D, Cortellessa G, et al. Engagement during interaction with assistive robots. International Journal of Social Robotics. 2018;**10**: 333-345

[53] Yun WH, Lee D, Park C, Kim J, Kim J. Automatic recognition of children engagement from facial video using convolutional neural networks. IEEE Transactions on Affective Computing. 2020;**11**(4):696-707

[54] Whitehill J, Serpell Z, Lin YC, Foster A, Movellan JR. The faces of engagement: Automatic recognition of student engagement from facial expressions. IEEE Transactions on Affective Computing. 2014;**5**(1):86-98

[55] Perry GT, Krippendorff K. On the reliability of identifying design moves in

protocol analysis. Design Studies. 2013; **34**(5):612-635

[56] Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960;**20**(1): 37-46

[57] Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;**16**(3):297-334

[58] Monarch RM. Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-Centered AI. New York, USA: Simon and Schuster; 2021

[59] Hemphill JF. Interpreting the magnitudes of correlation coefficients. The American Psychologist. 2003;**58**(1): 78-80

[60] Larson S, Mahendran A, Peper JJ, Clarke C, Lee A, Hill P, Kummerfeld JK, Leach K, Laurenzano MA, Tang L et al. An evaluation dataset for intent classification and out-of-scope prediction. arXiv preprint arXiv: 1909.02027. 2019

[61] Ljunglöf P, Larsson S, Mühlenbock KH, Thunberg G. TRIK: A talking and drawing robot for children with communication disabilities. In: Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009). Uppsala, Sweden: Northern European Association for Language Technology (NEALT); 2009. pp. 275-278

[62] Li J. Recent advances in end-to-end automatic speech recognition. Asia-Pacific Signal and Information Processing Association. 2022;**11**(1):e8

[63] Adipat S, Laksana K, Busayanon K, Asawasowan A, Adipat B. Engaging students in the learning process with game-based learning: The fundamental concepts. International Journal of Technology in Education (IJTE). 2021; **4**(3):542-552

[64] Axelsson A, Skantze G. Multimodal user feedback during adaptive robot-human presentations. Frontiers in Computer Science. 2022;**3**:55-68

[65] Javed H. Personalizing mixed initiative dance interactions with a socially-aware robot (doctoral dissertation, The George Washington University)

[66] Vachharajani P. Enhancing social media personalization: dynamic user profile embeddings and multimodal contextual analysis using transformer models. arXiv preprint arXiv: 2407.07925. 2024

[67] Anderson DR, Bryant J, Wilder A, Santomero A, Williams M, Crawley AM. Researching blue's clues: Viewing behavior and impact. Media Psychology. 2000;**2**(2):179-194

[68] Komatsubara T, Shiomi M, Kanda T, Ishiguro H, Hagita N. Can a social robot help children's understanding of science in classrooms? In: Proceedings of the Second International Conference on Human-Agent Interaction. New York, USA: Association for Computing Machinery (ACM); 2014. pp. 83-90

[69] Pop CA, Simut RE, Pintea S, Saldien J, Rusu AS, Vanderfaeillie J, et al. Social robots vs. computer display: Does the way social stories are delivered make a difference for their effectiveness on ASD children? Journal of Educational Computing Research. 2013;**49**(3): 381-401

[70] Bainbridge WA, Hart JW, Kim ES, Scassellati B. The benefits of interactions with physically present robots over

video-displayed agents. International Journal of Social Robotics. 2011;**3**(1): 41-52

[71] Alvarez-Benito G, Suarez C, Marquez C. Hello I am a robot and my name is Curro: Design, development and evaluation of a robot solution for linguistic, cognitive and physical rehabilitation of children after cancer treatment. PanCare Conference. 2019

[72] Neugnot-Cerioli M, Laurenty OM. The future of child development in the AI era. Cross-disciplinary perspectives between AI and child development experts. arXiv preprint arXiv: 2405.19275. 2024