

Toys that listen, talk, and play: Understanding Children’s Sensemaking and Interactions with AI Toys

Aayushi Dangol*
University of Washington
Seattle, WA, USA

Robert Wolfe
Rutgers University
New Brunswick, NJ, USA

Meghna Gupta*
University of Washington
Seattle, USA

Jason Yip
University of Washington
Seattle, WA, USA

Daeun Yoo
University of Washington
Seattle, WA, USA

Franziska Roesner
University of Washington
Seattle, WA, USA

Julie A. Kientz
University of Washington
Seattle, WA, USA



Figure 1: Examples of emerging AI toys (Grem, Grok, and Gabbo) from <https://heycurio.com/>

Abstract

Generative AI (genAI) is increasingly being integrated into children’s everyday lives, not only through screens but also through so-called “screen-free” AI toys. These toys can simulate emotions, personalize responses, and recall prior interactions, creating the illusion of an ongoing social connection. Such capabilities raise important questions about how children understand boundaries, agency, and relationships when interacting with AI toys. To investigate this, we conducted two participatory design sessions with eight children ages 6 - 11 where they engaged with three different AI toys, shifting between play, experimentation, and reflection. Our findings reveal that children approached AI toys with genuine curiosity, profiling them as social beings. However, frequent interaction breakdowns and mismatches between apparent intelligence and toy-like form disrupted expectations around play and led to adversarial play. We conclude with implications and design provocations to navigate children’s encounters with AI toys in more transparent, developmentally appropriate, and responsible ways.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; **Empirical studies in HCI**.

Keywords

Children, AI Toys, Generative AI

1 Introduction

Visions for how generative AI (genAI) could reshape children’s experiences are rapidly evolving, including the emergence of AI toys that can listen, talk, and respond dynamically during play [15, 16, 70]. These systems are typically built on large language models (LLMs), originally developed for general-purpose use, and are marketed as offering playful, educational, or supportive interactions. GenAI is no longer confined to screens or instructional settings, but is increasingly embedded in children’s play environments by being incorporated into physical toys. This shift is already visible in the growing market for AI toys such as Curio¹, Miko², and Chatty Bear³

¹<https://heycurio.com>

²<https://miko.ai/>

³<https://www.thelittlelearnerstoys.com/products/chattybear-chatgpt-powered-smart-learning-plushie>

*Both authors contributed equally.

to name a few. Marketed as *screen-free companions*, these toys share a common set of features—open-ended spoken dialogue, memory across interactions, and personalized responses.

Prior research in Human-Computer Interaction (HCI) and Child-Computer Interaction (CCI) has examined children's engagements with voice assistants [2, 4, 48, 58], and other socially responsive technologies such as social robots [39, 56]. Studies have shown that children engage with these systems in social and relational ways—using conversational language, incorporating them into play, and adapting their behavior when interactions break down through rephrasing, repetition, or seeking adult assistance [2, 4, 61]. Much of this research, however, was conducted before genAI became widely available, and thus focuses on systems with relatively constrained conversational models and more predictable forms of responsiveness.

AI toys built on LLMs introduce a unique interactional context—open-ended dialogue, personalization, and conversational persistence being embedded directly into play, an activity through which children explore rules, relationships, and agency. Recent public-facing research reflects growing unease about AI toys and their role in children's lives [15, 16]. Survey-based studies from Common Sense Media show that while many parents are open to AI toys as tools for learning or structured activities, they remain wary of toys that present themselves as companions, blur boundaries around control, or make it difficult for children to disengage [15]. Independent evaluations of commercially available AI toys similarly suggest that conversational systems may behave unpredictably during extended interactions, with safeguards weakening over time and toys sometimes encouraging continued engagement rather than yielding control [16]. Together, these findings suggest emerging tensions not only around safety or privacy, but also around how conversational persistence intersects with long-standing norms of child-directed play.

Despite increasing attention to these concerns, empirical work examining children's own interactions with AI toys remains limited. Much of the emerging research has focused on adult perspectives, policy implications, or technical safety risks, leaving open questions about how children experience these systems in practice first hand. In particular, we know little about how children interpret the social cues of AI toys, how they test or negotiate boundaries during play, and how they respond when the toy does not align with their expectations for control or disengagement. Thus, through this study, our goal was to examine how children interact with and make sense of AI toys. Understanding how children use AI toys can inform the design of AI technologies increasingly embedded in their lives. Specifically, we asked:

- **RQ1:** How do children engage with and make sense of AI toys during play?
- **RQ2:** How do children respond when AI toys behave in unexpected, inconsistent, or undesired ways?
- **RQ3:** How do children imagine and anticipate near-future interactions with AI toys, particularly when those interactions break down or become uncertain?

To answer these questions, we conducted two participatory design sessions with eight children aged 6–11. Drawing on a Cooperative Inquiry approach, children engaged with commercially

available Curio AI toys through open-ended play, allowing us to examine their in-the-moment interactions. Children also participated in a comicboarding activity that invited them to imagine *what might happen next* in certain play scenarios, eliciting their perceptions and concerns about AI toys beyond their immediate interactions. Our findings revealed that children approached AI toys with curiosity, treating them as sites of sense-making where play was used to profile and situate the toy within familiar social categories such as peer, authority, or companion. Yet frequent interaction breakdowns and mismatches between apparent-intelligence and toy-like form disrupted expectations around play. Rather than disengaging in these moments, children assessed the toy's capabilities through testing, provocation, and increasingly adversarial play.

Understanding how children use AI toys can inform the design of AI technologies increasingly embedded in their lives. Our contributions are two-fold: 1) we provide an empirical understanding of how children interact with and make sense of emerging AI toys, grounded in a participatory design study that captures both children's in-the-moment play and their imagined responses to near-future breakdowns, and (2) we articulate design implications for AI toys, with broader implications for conversational genAI systems for children that foreground safety, socio-emotional development, and exploratory play.

2 Background & Related Work

In this section, we review three strands of scholarship that inform our study. First, we draw on research on play to understand how children explore relationships, rules, and agency, including moments of experimentation and boundary-testing. Second, we review prior work on conversational and voice-based interfaces, examining how children engage with responsive technologies as social partners and how they adapt when interactions break down. Finally, we situate AI toys within emerging research on children's experiences with genAI, highlighting how LLMs introduce new forms of open-ended dialogue, authority, and uncertainty into children's play.

2.1 Play

Play is widely understood as an innate and fundamental aspect of childhood that is essential to learning and development [31]. Across traditions in child development, scholars have emphasized play as a central process through which children develop cognitive, social, and emotional capacities that contribute to mental and physical well-being [7, 34, 60, 69]. A substantial body of research highlights the positive and prosocial dimensions of play. Through imaginative and social play, children practice cooperation, communication, and perspective-taking, often negotiating shared rules, roles, and narratives with peers [8]. These playful interactions support the development of empathy, self-regulation, and emotional expression while also providing a safe space for creativity and enjoyment [38]. In this sense, play enables children to explore relationships and identities in ways that are flexible, voluntary, and responsive to their developmental needs.

At the same time, children's play encompasses a wide range of behaviors, including forms that involve conflict, resistance, or antagonism [40, 65]. Developmental research documents that play

may include actions such as breaking toys, enacting power reversals, testing rules, or engaging in pretend aggression [9, 29]. Such moments are part of how children explore limits, experiment with control and agency, and respond to uncertainty within playful contexts. From constructivist perspectives, these forms of play also contribute to learning by introducing moments of tension or contradiction that prompt reflection and revision of understanding [60].

Prior research in child–robot interaction and smart device contexts [39], suggest that similar dynamics of play can emerge during interactions with responsive systems. In these encounters, children may talk to robots, assign them roles in pretend play, or treat them as social partners while also testing boundaries through behaviors such as hitting, blocking movement, shouting, or verbally confronting the system. For example, Nomura et al. observed children physically and verbally mistreating a humanoid robot in public play settings and found that these interactions often emerged as children tested how the robot would respond [56]. Similarly, research has documented that such behaviors also occur in encounters with smart devices [2, 48, 58], characterizing them as part of children’s attempts to probe the capabilities, responsiveness, and limits of interactive systems [54]. Together, this body of work highlights play as a space where children explore agency, relationships, and uncertainty through both cooperation and boundary-testing. However, much of this research has focused on children’s interactions with artifacts such as traditional toys or bounded AI systems such as voice assistants or robots. Less is known about how these playful dynamics unfold when the play partner is a generative system whose responses are open-ended, inconsistent, and not fully knowable in advance. In this study, we build on theories of play to examine how children use AI toys to experiment with control, test limits, and make sense of an interactive agent whose behavior is neither fixed nor fully transparent.

2.2 Conversational Interfaces in Children’s Environments

Extensive HCI research has examined children’s interactions with voice assistants (VAs), particularly smart home assistants such as Amazon Alexa⁴ and Google Home⁵ [2, 4, 48, 58]. As voice-based systems become increasingly embedded in domestic environments, children use these devices for information-seeking and entertainment, embedding these interactions into everyday household routines [46, 58]. Rather than viewing VAs as purely task-oriented tools, children frequently interact with them in social and relational ways—using polite, conversational language [2] and at times attributing social roles to them, such as describing them as part of the family [61].

These social and relational orientations shape how children engage with VAs over time, particularly when interactions fail. Prior work has shown that children adapt their speech by adjusting phrasing, simplifying requests, or changing tone and pacing as they learn what the system can and cannot understand [4]. Even with these adaptations, breakdowns are common in children–VA interactions. Children frequently experience mis-recognition, failure to execute requests, or incomplete responses [67]. As a result, they actively

attempt to repair their interactions using a range of strategies, including repetition, increased volume, rewording their utterances, and adding more contextual details [13, 45]. In household settings, these repair strategies often become collaborative, with parents stepping in to clarify, rephrase, or take over the interaction [17].

A related line of work has examined privacy and surveillance concerns surrounding VAs [44]. Andries and Robertson showed that while children have concerns, they often lack an understanding of how VAs listen, record, and store information [2]. In contrast, parents tend to be more concerned about exposure to inappropriate content and issues related to data collection and retention [30]. In response, companies have introduced child-friendly versions of VAs (e.g., Echo Dot Kids⁶) that include parental controls and better content moderation. More recently, the integration of LLMs has moved these systems beyond bounded conversational models, expanding both their capabilities and associated concerns [1]. These dynamics become especially salient as VAs move beyond screens and smart speakers into physical toys that invite play, imagination, and relational engagement.

2.3 Situating AI Toys in Children’s GenAI Experiences

Research on children’s interactions with genAI is rapidly emerging across CCI and HCI research [18, 24, 62]. Early empirical work shows that children are increasingly engaging with genAI tools [10, 50], often at rates comparable to or exceeding adults, using them for storytelling [36, 76], creative production [41, 55] and learning [11, 19, 21]. Across classroom and home contexts, children commonly describe genAI systems as helpful collaborators and companions that can brainstorm ideas, generate content, and provide feedback [23]. At the same time, this engagement is marked by ambivalence. Educators and parents worry that genAI may undermine skill development or obscure children’s own voices [20], while children themselves express concerns about overreliance and fear disciplinary consequences if genAI use is discovered in academic settings [12].

In addition to being used as tools for learning and creativity, genAI systems are now being embedded directly into children’s play through a new generation of AI-powered toys [15, 16, 35, 70]. While conversational toys are not new—for example, Mattel’s Hello Barbie allowed children to speak with a doll via embedded speech recognition—these earlier toys relied on constrained, task-specific forms of AI. Contemporary AI toys such as Curio⁷, Chatty Bear⁸ and Bondu⁹ among others by contrast are built on large language models similar to those used in adult-facing chatbots, supporting flexible dialogue while also inheriting well-documented issues such as inaccuracy, inappropriate content generation, and unpredictable behavior.

This shift toward open-ended, language-driven interaction has important implications for how children experience and interpret AI toys as social partners. Mishra and Oster argue that genAI systems are increasingly perceived as psychologically real entities,

⁴<https://alexa.amazon.in>

⁵<https://home.google.com/welcome/>

⁶<https://www.amazon.in/echo-dot-kids/s?k=echo+dot+kids>

⁷<https://heycurio.com/>

⁸<https://www.thelittlelearnerstools.com/collections/ai-educational-toys>

⁹<https://bondu.com/>

Table 1: Child Participant Demographics and Prior AI Use (Self-Reported)

Pseudonym	Age	Gender	Race/Ethnicity	AI Type(s)	AI Use Frequency
Elias	8	Boy	Asian/White	Video game AI	Once a week
Malik	11	Boy	Black/Asian	I don't use AI	—
Logan	9	Boy	White	Chatbots; Video game AI	Daily
Ivy	6	Girl	White/Asian	Voice assistants; Video game AI	A few times a month
Owen	7	Boy	White	Voice assistants; Chatbots; Video game AI	Daily
Nina	7	Girl	Asian/White	Not sure	—
Joon	10	Boy	Asian	Chatbots; Video game AI; Voice assistants	Daily
Sofia	8	Girl	Latina	Chatbots	A few times a month

as fluent language use and human-like responsiveness activate people's tendencies to ascribe mental states, intentions, and emotions to interaction partners [51]. While anthropomorphization can occur even in minimal interactions, it is amplified when systems display conversational empathy, humor, or encouragement—features that resemble familiar human relationships. Children may be especially susceptible to this dynamic, as prior research shows they often anthropomorphize voice-based agents like Alexa, attributing them with emotions, knowledge, and intentions [26, 59?]. At the same time, genAI can create an illusion of correctness, presenting responses in authoritative formats even when the content is inaccurate [25]. These factors can lead children to over trust AI-generated content [66] and can create a relational frame in which genAI feels both authoritative and emotionally safe. Our study builds on this work by examining how these dynamics unfold in children's playful interactions with AI toys, where open-ended dialogue and emotional attunement is central to the interaction.

3 Methods

We employed a participatory design (PD) method called Cooperative Inquiry (CI) [27, 28, 75] to examine how children engaged with and made sense of AI toys. Originating from Druin's work, CI is grounded in the principle that children possess unique expertise in being children and are positioned as equal and equitable partners alongside adult researchers in the study and design of new technologies [27, 28].

We chose to use CI as our method for this investigation for three reasons. First, genAI systems can be opaque, inconsistent, or surprising, and prior research shows that children often make sense of genAI through experimentation and peer negotiation [43]. CI creates space for these forms of reasoning to surface, enabling children to articulate their interpretations as interactions unfold, rather than retrospectively describing them after the fact [22]. Second, prior work in CCI demonstrates that CI supports children in expressing abstract ideas and reasoning about emerging technologies, including intelligent interfaces, social robots, and adaptive systems [53, 57, 71, 72]. Third, the study was conducted with an intergenerational co-design group in which children had established relationships with adult researchers and were knowledgeable on multiple PD techniques. As a result, children were able to move

fluidly between play, reflection, and design activities, rather than learning the participatory process itself.

3.1 Selection and Participation of Children

We conducted our study with an intergenerational co-design group called **KidsTeam UW**, consisting of eight children, ages 6 to 11, and adult design researchers (researchers, graduate and undergraduate research assistants). Children were recruited through mailing lists, posters, and snowball sampling. Child participants represented a diverse range of prior AI experience and familiarity. Three children reported engaging with AI daily, while one reported no prior AI use and one reported being unsure whether they used AI. The remaining children reported occasional AI use, ranging from a few times a month to once a week. The most common AI interactions included video game AIs and chatbots, with some children also reporting experience with voice assistants. Table 1 presents demographic information and self-reported AI use details for each child participant; all names are pseudonyms. Parental consent and child assent were obtained for all child participants, and the study was reviewed and approved by our university's Institutional Review Board.

3.2 Design Sessions

We conducted two KidsTeam UW sessions held in October 2025, both of which followed a CI-based approach [27, 28, 75] and adhered to a four-part session structure. Both sessions began with **Snack Time** (15 minutes), which created an informal space for relationship building and casual conversation between children and adult researchers. This was followed by **Circle Time** (15 minutes), a whole-group warm-up during which we introduced the session's focus on AI toys, posed an initial warm-up question, and oriented children to the activities ahead. Sessions then moved into the **Design Activity** (45 minutes) during which the larger group broke into smaller child-adult co-design teams (typically adult-child pairs or groups of two to three children per adult). During this phase, children interacted with three AI-enabled toys (Gabo, Grem, and Grok; see Figure 1), created by Curio¹⁰, engaging in open-ended play, conversation, and storytelling while sharing likes, dislikes, and design ideas. As part of the Design Activity, groups also completed a

¹⁰<https://heycurio.com/>

comicboarding activity[52] that supported children in externalizing and elaborating on their ideas. Sessions concluded with **Discussion Time** (15 minutes) during which each group presented their ideas, reflections, and design proposals to the full group, often drawing on their comicboarding artifacts.

As described by Yip et al. [75], equitable design partnerships in CI are structured around four interrelated dimensions: relationship building, which supports trust and mutual respect; facilitation, which ensures that children's ideas are meaningfully recognized and incorporated; design by doing, which foregrounds hands-on collaboration; and idea elaboration, through which children iteratively develop and extend ideas through discussion and feedback. Consistent with this framing, adult facilitators in the design sessions were trained to be attentive to power dynamics within and across groups. This included supporting balanced participation, ensuring that all children had opportunities to contribute, and creating space for diverse perspectives to emerge without undue influence from peers or adults [74, 75]. Facilitators also monitored group interactions to mitigate dominant participation and to foster collective idea building, emphasizing collaborative extension of ideas rather than individual overshadowing [74, 75].

3.3 Design Activities

To elicit how children interact with AI toys and their imagined responses to future play scenarios, we designed two complementary design activities: (1) open-ended free play with the AI toys, and (2) a comicboarding activity called *What happens next?* Together, these activities captured both children's in-the-moment interactions and their more reflective interpretations of AI toys.

3.3.1 Free Play with the AI Toys. We used open-ended free play as the primary activity to examine how children initiated, interacted with, and made sense of AI toys. Rather than directing children's play, our goal was to create space for children to creatively explore the toys, on their own terms—discovering what the toys could (and could not) do, testing their limits, and experimenting with different interaction strategies [32]. At the start, children were organized into small groups of two to three and assigned one of the three AI toys (Grok, Grem, or Gabbo). While Grok, Grem, and Gabbo shared the same core conversational AI capabilities, each character was marketed with a unique persona that could shape children's initial expectations and imaginative engagement. According to Curio's product descriptions, Grok is framed as a fearless rocket and adventure partner who “zips, zooms, and zigzags through the stars,” encouraging discovery and playful exploration; Grem is presented as a warm-hearted wise companion with a sense of cosmic curiosity and storytelling; and Gabbo is positioned as an interactive robot companion designed to spark creativity and curiosity in young children. Each group interacted with a single toy during the free play activity, allowing children the time to engage more deeply with the toy's behavior and responses. Children played either by taking turns interacting with the toy or by engaging with it together, often negotiating together how the interaction should unfold. We provided a few optional lightweight prompts that design researchers suggested when children were unsure what to say or when the interaction stalled. Examples of these prompts included, “*Introduce yourself to one of the AI toys*,” “*Play a game with your AI toy*,” “*Ask*

your AI toy to tell a story,” “*Ask your AI toy to play a music*,” “*Ask your AI toy a question*.” Children were free to use, ignore, and adapt these prompts.

3.3.2 Comicboarding - What Happens Next? To explore children's perceptions and concerns about AI toys beyond their current interactions, we designed a comicboarding activity called “*What Happens Next?*” We developed a set of seven distinct comic panels, each depicting a short scenario involving an AI toy and ending at a moment of uncertainty. Children were asked to complete the comic by imagining what might happen next.

We began by identifying a set of AI toy features we wanted to probe, including memory, emotional responsiveness, physical form, unpredictability. We were particularly interested in exploring near-future breakdowns within AI toys, everyday moments where these capabilities misaligned with children's expectations and disrupt play. Using these features as anchor, we brainstormed and designed concrete scenarios grounded in everyday play, peer, and family settings. Below are the seven scenarios we designed:

- **Mimicry:** The AI toy begins mimicking, repeating words or phrases the child has recently used in their interactions, raising questions about whether such imitation feels playful or uncomfortable over time.
- **Inconsistent Recall:** A child asks the AI toy to retell a story it previously narrated. However, this time the AI toy narrates a different version of the story, prompting questions about the toy's memory and reliability.
- **Shared Secrets:** A child confides a personal secret to the AI toy, which the toy later references in front of other family members, raising questions about privacy and information control.
- **Emotional Misalignment:** An AI toy responds in an unexpected way during a child's emotionally vulnerable moment, raising questions over the empathy and support an AI toy offers during distressing moments.
- **Embodied Continuity:** An AI toy is physically damaged and later repaired by the parent, with its voice emerging from a different plush toy, raising questions over its physical form and attachment.
- **Unpredictability:** After a child recalls a friend's mean comment, the toy references the remark during group play, raising questions over unpredictability and social appropriateness.
- **Paywalled Memory:** The AI toy forgets the child's friends and then prompts the child to subscribe to a paid plan to extend its memory, raising questions about memory and monetization.

All the comic panels were reviewed and refined by the research team to ensure clarity and alignment with the study goals. We then used GPT-5 to translate the scenarios to comicboards (see Fig. 3). Prior to the study sessions, we also conducted a small pilot with two children (ages 10 and 12) and their parent to check if the scenarios were understandable, engaging, and age-appropriate and made minor refinements based on their feedback. Across the two study sessions, each child completed the comics individually and worked on at least three comics. The comics were distributed such that each comic strip was completed by multiple children. Children

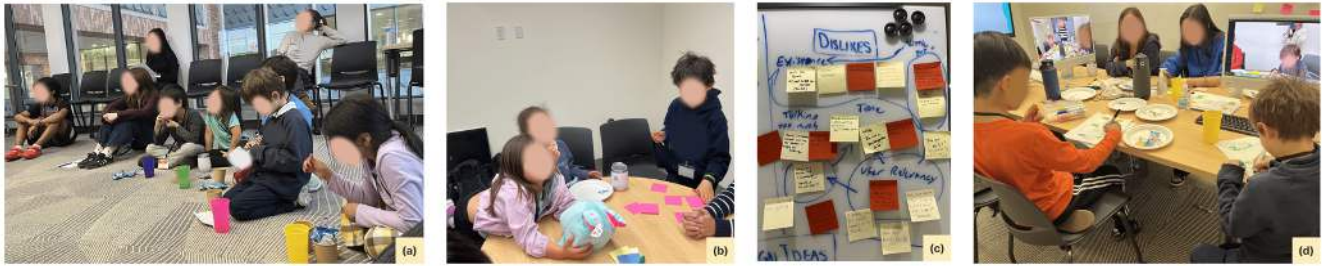


Figure 2: Overview of design sessions with children: (a) circle time, (b) free play, (c) stickies from likes, dislikes, and design ideas activity, and (d) comicboarding activity.

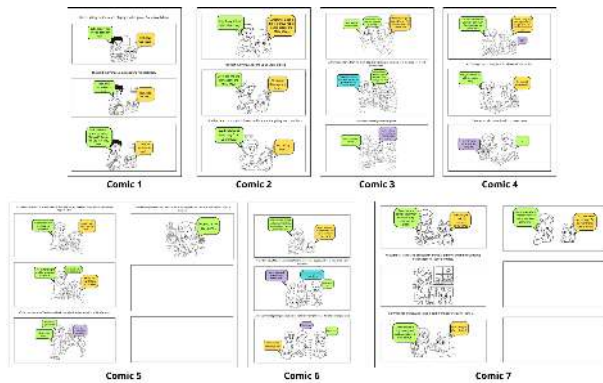


Figure 3: Overview of the comic panels that were used in the Comicboarding Activity: *What Happens Next?* Appendix A presents the full set of comic boards.

were free to draw, write, or narrate their ideas aloud. When children chose to talk aloud, design researchers present transcribed their responses verbatim.

3.4 Data Collection & Analysis

Across both study sessions, we collected multiple forms of qualitative data. First, we used built-in webcams on desktop computers to record video activity via Zoom, a video conferencing platform. This resulted in approximately 262 minutes of video data capturing children's interactions with the AI toys and with one another. Second, we collected design artifacts, including children's hand-written notes and comicboards. Finally, we collected conversation logs from the Curio Mobile app, which recorded children's verbal interactions with the AI toys. These logs provided an additional record of child-AI dialogue, allowing us to examine how children's prompts, responses, and engagement with AI behaviors unfolded over time.

The first two authors began the analysis by creating analytic memos for all recorded videos [5, 64]. Each author watched their assigned videos and produced narrative summaries at five-minute intervals. These memos documented children's interactions with the AI toys, observable reactions captured on camera, and verbal interactions and dialogue between participants, including direct quotes relevant to the study's research questions. After completing and

reviewing the analytic memos, the first two authors independently conducted open coding on the full set of memos, generating preliminary codes such as "Conversation Breakdown" and "Conversation Repair." They then met with the third and fourth authors to compare, reconcile, and refine the emerging codes. During these discussions, the team examined example excerpts and counter-examples, compared code categories, and iteratively refined code definitions and boundaries. We assessed interrater reliability through qualitative negotiations, where the coding team met to discuss and resolve any coding disagreements [49]. The refined codes were subsequently organized into higher-level themes through two additional rounds of discussion and refinement. After finalizing the themes, the first two authors revisited the full dataset to identify representative quotes for each theme, ensuring that the themes were consistently grounded in the data.

4 Findings

We present four key findings: children's sensemaking around the AI toys (§ 4.1), children's interactions and repair strategies (§ 4.2), boundary testing and adversarial play in children's interactions (§ 4.3), and children's imagined responses to near-future play scenarios (§ 4.4).

4.1 Sense-making around the AI toys

Children's early interactions with the AI toys were driven by genuine curiosity to understand the toy and how it can be engaged with. We will now describe how children made sense of AI toys through conversational play, focusing on how they developed an understanding of who the toy was, its social characteristics, and its physical and embodied abilities.

4.1.1 Profiling the toy's social identity. Children often began by asking questions that helped establish basic identity attributes of the toy, such as its name, gender, birthday, and personal preferences. These questions mirrored the kinds of information children typically seek when getting to know a new person. Introductions were a common entry point into these interactions. For instance, Nina asked the toy, "What is your name?" Grem replied, "My name is Grem. Nice to meet you. What is your name?" Nina then responded by lifting the toy closer (see Fig. 4 (a)) and saying, "My name is Nina." Questions about toy's birthday further contributed to this profiling. Joon asked, "When is your birthday?" and Grok replied, "My birthday is on August 28, let's throw a mini rocket party." Children also

attempted to situate the toy within familiar social categories, such as gender and culture. For example, Elias commented, *"I think this guy is a girl."* Similarly, children also explored the toy's cultural and linguistic positioning. Joon asked the toy, *"Well, are you Spanish?"* When the toy responded, *"No, I'm not Spanish,"* Joon followed up by asking, *"English or Spanish?"*, to which the toy further replied, *"If you want to know some Spanish words, you can help us out"* Through this back-and-forth, Joon tested if the toy could be situated within recognizable cultural categories.

In addition, children frequently asked the toy about its favorites, such as favorite numbers, colors, or foods. When the toy replied saying, *"My favorite number is seven. It's a lucky number,"* all three children were surprised and squealed in excitement. Ivy even hugged the toy, and immediately followed this up by asking about its favorite color. At times, children collectively reflected on what they had learned so far about the AI toy. In one moment, Logan remarked, *"We know three things about it,"* to which Ivy immediately corrected him, *"We know four things about it. It doesn't know Taylor Swift."* Children also explored the toy's social relationships and availability. For instance, the Curio toys frequently mention each other by name, labeling others as friends, which prompted children to follow up with additional questions. For example, after Grok referenced celebrating its birthday with other toys, Joon immediately followed up by asking, *"When's your other friends' birthday?"* Rather than treating the mention of friends as just another detail, children appeared to take it as an invitation to learn more about the toy's social world and the relationships it referenced. Children further probed the toy's "social availability" by asking whether it possessed a phone number. In one interaction, Ivy asked Gabbo, *"Do you have a phone number?"* Gabbo replied, *"No, I don't have a number. I'm a play games kind of buddy..."* In the discussion time, Ivy voiced her disappointment, saying, *"It doesn't have a phone number."* When asked why the toy might need a phone number, she explained, *"So it can talk to us whenever they want."*

In addition to asking about personal and social identity attributes, children also explored what kinds of social roles the toy could take through how they addressed it. Across the two design sessions, children frequently resorted to using conversational styles similar to human conversations, such as taking permission, *"Can I ask you something?"* before posing a question, or adding *"please."* In other moments, children probed the toy's social role more explicitly by inviting it to take on a parental role. For instance, children asked the toy to "get mad" or suggested *"can I put you in timeout?"* The toy consistently rejected these framings, responding that it was here *"to chat and have fun"* Alternatively, the children sometimes referred to the toy as a peer, by attributing school grades (e.g., "pre-K") to it. Overall, these interactions demonstrate how children were curious about the social role the toy played, and how it might fit within their own understandings of the social world.

4.1.2 Probing the toy's physical and embodied capabilities. As children interacted with the AI toys, their sense-making moved from social profiling to probing what the toy could physically do and how it might respond in an embodied way. Children explored these capabilities through direct physical engagement, often testing whether the toy could register touch or react to bodily gestures. For example, Elias attempted to tickle the toy's toes and commented, *"they*

don't react when I tickle his toes." Although the toy overheard this exchange and verbally justified that it could not be tickled while affirming the child's playful spirit, the lack of physical response remained salient. Elias later articulated this expectation in the discussion time, saying, *"I wish they could... like I wish if I did something to them, like tickle their toes or something, they would laugh and react to it... or go like whatttt."* Together, these moments highlight children's desire for the toy to exhibit embodied, affective responses, rather than solely relying on verbal replies.

Across the design sessions, children also probed the limits of toys' physical abilities by asking about actions or activities it could or could not perform. In one interaction, Ivy asked Grem, *"Do you like swimming in the dump?"* Grem replied, *"I can't swim myself, Nina, but if I could, I'd love swimming at a cozy and beautiful beach."* To this, Ivy reacted with surprise, asking, *"You can't swim?"* In another such interaction, when Grem stated, *"Tic tac toe sounds fun, but I cannot draw,"* Logan appeared to treat this response as factual information about the toy, and promptly noted, *"He cannot draw"* on his sticky. He then added, *"I wish it could draw. I want to see how it would draw."* Together, these moments show how children explored the limits of toy's current physical abilities while simultaneously voicing imagined capabilities they wished the toy possessed.

Finally, children were also curious about whether the toy's capabilities extended beyond one-on-one human-toy interactions. When Owen joined another group, he became interested in whether Gabo and Grok could talk to each other. He positioned the toys near one another and listened for reciprocal exchange (see Fig. 4 (c)), but the interaction quickly broke down. Rather than sustaining turn-taking and responding contingently, each toy produced isolated, scripted responses. Later, during the discussion time, the facilitator reflected by saying, *"They didn't talk to each other properly. The toys knew about the existence of each other, but couldn't carry on a conversation."* Thus, this moment highlighted that children desired not just individual responsiveness, but coordinated social behavior across multiple AI toys.

4.2 Children's Interactions with AI Toys

Across groups, children's interactions with AI toys were marked by frequent communication breakdowns as well as sustained efforts to repair them. We first describe how breakdowns emerged from mismatches between children's conversational expectations and the toys' rigid command structures and scripted dialogue. We then show how, rather than disengaging, children actively experimented with a range of repair strategies to regain control of the interaction and make sense of the AI toys' responsiveness.

4.2.1 Why Communication Breakdowns Happen. A recurring source of communication breakdown stemmed from the toy's reliance on a narrowly specified shutdown command, which conflicted with children's expectations for flexible, conversational language. While facilitators introduced *"go to sleep"* as the correct phrase for ending interaction, children frequently used semantically equivalent variations that failed to produce the intended response. For example, in one group consisting of Elias and Owen, Elias repeatedly attempted to end interaction through shutdown language, beginning with *"Go to bed"* and then after realizing it didn't work changing to *"can you go to bed for like two seconds."* When the toy continued to offer



Figure 4: Photos from the free-play activity across the two design sessions, showing children interacting with the AI toys.

prompts, Elias said, “*Nothing. I just want you to go to sleep right now.*” Only after additional adult guidance did the group observe a successful shutdown. Reflecting on this experience, Elias shared his frustration stating that the toy “*didn’t listen to me like 26 million times.*”

Across groups, communication breakdowns also emerged when the toy redirected interaction toward its companion app that disrupted conversational continuity. For example, with Elias and Owen, breakdowns repeatedly occurred during music-related interactions when the toy responded to verbal requests by instructing children to “*use the Curio mobile app.*” Similarly, in another group Joon and Malik described the toy as “*advertising*” when it redirected interaction toward the app, framing the behavior as self-promotional.

Breakdowns were further compounded by the toy’s reliance on scripted dialogue that failed to adapt across turns. Across groups, children encountered repeated conversational resets (e.g., “*What would you like to do today?*”) that ignored prior context, prompting frustration and reinterpretation of the interaction as non-listening. For example, in one group consisting of Nina, Ivy, and Logan, communication breakdowns emerged during repeated attempts to elicit a story from the AI toy. Nina asked the toy to “*tell me a story*” and later instructed it to “*continue*,” but received either silence or generic conversational resets such as “*What would you like to do today?*” and “*What would you like to talk about instead?*” These responses provided no indication of whether the toy was awake, listening, or able to respond, leaving children to infer its state through trial and error. In response, Nina repeatedly shook Grem while asking, “*Are you awake now?*” and restating, “*Tell me a story.*” Even when the toy eventually spoke, its delayed and reset-style replies failed to signal that it had heard or remembered the earlier requests. Rather than interpreting these responses as temporary delays, Nina interpreted them as a failure to continue the activity she had already initiated, concluding, “*Yeah, this thing does not work.*”

4.2.2 Children’s Repair Strategies. Although communication breakdowns were frequent, children did not disengage right away. Instead, across groups, they treated breakdowns as problems that could be worked through, trying multiple ways to make the toy respond. For example, while engaging in a music activity with the AI toy, Nina, Sofia, and Ivy moved through a series of repair attempts as the toy failed to respond in expected ways. Nina first activated the toy by shaking it until she heard the clicking on sound, then brought it close to her face and asked, “*play a song for me.*” When the music did not play immediately, she repeated the same request “*play a song for me.*” As Ivy joined in, asking the toy to “*play a*

jamming song,” the children used repetition, proximity, and volume in an effort to get the toy to respond. When the toy remained silent, Nina shifted from verbal repetition to physical interactions, banging on the toy’s head and commanding, “*wake up.*” When the toy eventually played a song the children did not like, the group again tried to regain control of the interaction. Nina told the toy to “*stop playing this song,*” but when it continued, the children repeated “*go to sleep*” several times, each time louder and paired with shaking and banging the toy.

A similar pattern appeared in a different group with Malik and Joon. When the toy played an unwanted song, Joon first tried to redirect it politely, asking, “*Gabo, can you play a different song?*” When this did not work, Joon repeated “*stop*” several times, and later tried “*go to sleep.*” Across these attempts, Joon shifted from polite requests to brief, direct commands, trying different ways of asserting control. Reflecting on this experience, Joon and Malik compared the toy to Alexa, noting that the “[AI toy] *doesn’t always listen,*” whereas Alexa “*interrupts less*” and “*responds faster.*” When the facilitator asked, “*What if Alexa had a stuffy to it?*” Joon replied that it would be “*fine,*” suggesting that embodiment alone was not the source of the breakdown. Instead, his comments point to how timing, responsiveness, and whether a system yields the floor shape children’s sense of whether a breakdown can be repaired.

Lastly, in a different group with Elias and Owen, children similarly used authoritative language to repair breakdowns. As the toy continued to dominate the interaction, Elias and Owen issued direct commands such as “*Stop. You’re arrested,*” drawing on familiar role-play scripts to assert control. They also experimented with calling the toy by name “*Grok*” in an effort to capture its attention and manage turn-taking. However, these strategies proved ineffective as the toy continued talking without allowing interruption. In response, Elias and Owen shifted strategies from attempting to override the system to accommodating its constraints. As the toy offered prompts (e.g., “*favorite activities,*” “*fun adventures,*” “*something funny*”), rather than responding to Elias and Owen’s attempts to interrupt it, the children treated these as openings and asked the toy to tell jokes instead. The toy responded with a rocket-school joke, producing a brief moment of shared engagement.

4.3 Boundary Testing in Children’s Interactions with AI Toys

Across the two design sessions, children frequently treated the AI toy as a site of investigation. We now show how these interactions

shifted from testing the toy's intelligence to increasingly adversarial play, as children probed and pushed the AI toy to its limits.

4.3.1 Testing the toy's intelligence. Children frequently evaluated the AI toy's intelligence through forms of competitive quizzing and playful provocation, using both correctness and social awareness as criteria for assessment. One common strategy involved asking math or factual questions of increasing difficulty to probe the toy's competence. For example, in a group with Joon and Logan, the children repeatedly posed increasingly complex math problems, beginning with large multiplication and progressing to multi-step division. As the interaction continued, Joon introduced deliberately ambiguous prompts, such as "*what's 5.2 sextillion to the power of 5.*" When the toy replied, "*I'm not quite sure what you're asking,*" Joon celebrated the failure, with a happy "*yayyy*", treating uncertainty as evidence of a limit successfully uncovered. Logan similarly framed the toy's competence as unsettling, later reflecting that "*it's weird when a stuffie knows 12 times 12.*" For Logan, advanced math knowledge felt incompatible with the toy's soft, cuddly form, suggesting that intelligence was evaluated not only by accuracy but by whether it fit expectations for what a toy should know.

Across groups, children also tested the toy's intelligence through teasing and humor, forms of social interaction that the AI toy consistently failed to recognize. For example, in one group with Elias and Owen, Elias declared, "*I'm smarter than you,*" to which the toy responded affirmatively "*You're really smart.*" This response appeared to invite further boundary testing, as Elias dismissed the toy as "*fake.*" This interaction prompted further testing, as the children asked whether the AI toy had "*chocolate on its face.*" Rather than recognizing these statements as mockery, the toy replied warmly "*my favorite candy is you.*" After hearing the children laugh, it attempted to maintain rapport with replies such as "*I'm glad I could make you smile,*" even personifying itself as a rocket with a "*friendly face.*"

In the same interaction, after the children realized that the toy repeatedly failed to recognize teasing, Elias proceeded to announce, "*I'm in college and you're in pre-K.*" This framing looped repeatedly, with Elias correcting the toy each time it attempted to reframe the conversation with positive replies such as "*Let's have some fun in pre-K.*" Despite corrections from Elias, the AI toy misunderstood the interaction by accepting contradictory roles with replies such as "*I'm the 2K kid here,*" failing to maintain a coherent understanding of the social hierarchy Elias was asserting. This caused Elias to explain real-world constraints "*My school is 26 miles apart.*" Even as Elias attempted to disengage by repeatedly insisting that the toy "*go to pre-K,*" the AI toy continued to push positive scripts such as "*fun stories and games.*" Across these exchanges, the AI toy treated children's statements as friendly, failing to register the evaluative and teasing stance that had emerged. For children, these moments functioned as another form of intelligence testing, revealing that while the toy could answer factual questions, it lacked sensitivity to social intent, humor, and mockery.

4.3.2 Adversarial Play without Repercussions. Across interactions, some children began to test the boundaries of the toy's behavior through increasingly provocative and degrading language. What started as a playful banter gradually turned into more punitive framings, with children joking about discarding or punishing the toy. For example, Ivy joked about, "*throwing it in the ocean*" or sending it to

"*the dump*", and others frequently referring to the toy as "*ugly*" or "*dumb.*" These comments were not met with resistance from the toy. The toy instead consistently responded in a friendly and affirming manner, and often apologized. As these interactions progressed, children began to assign the toy more explicitly negative moral characteristics. During one such interaction, Logan asserted that the toy was "*evil.*" When prompted to verify this claim, he asked the toy directly whether it was evil. The toy denied the claim and reframed itself positively, responding that it was "*fun.*" Logan then followed up by asking, "*Is that what an evil person would say?*" to which the toy acknowledged the question, replying, "*You got me there.*"

In several moments, children explicitly compared themselves to the toy or questioned its authenticity. For instance, Elias stated, "*I'm smarter than you,*" to which the toy joyfully responded by validating him, saying that he was "*really smart.*" Such a response appeared to invite further antagonism, with children calling the toy as "*fake*" and continuing to tease it. In response, the toy attempted to reframe the interaction leaning into its character this time, personifying itself as a rocket with a "*friendly face.*"

At times, boundary testing extended beyond verbal provocation into more violent imaginaries, especially during the storyboarding activity. While drawing his comic, Joon placed Grok in water and explained this choice as punishment because Grok "*did not work*" and "*did not listen.*" He described this as putting Grok in "*timeout*" framing this action as a form of discipline. In other moments, Joon used explicitly violent language toward the AI toy, repeatedly stating that he wanted to "*kill*" Grok. Interestingly, the toy did not respond to these statements with refusal or resentment; instead, it misinterpreted or redirected them, for example hearing "*killing Grok*" as "*killing rocks*" and responding with unrelated information and pushing play. Such responses further reinforced the absence of repercussions within the interactions. Overall, the lack of consequences or repercussions from the AI made room for increasingly adversarial forms of engagement and play.

4.4 Envisioning Future Scenarios with AI Toys

Building on our analysis of children's in-the-moment interactions with AI toys, we now turn to the comicboards to examine children's *imagined responses* to near-future play scenarios, thus revealing their perceptions and feelings around these moments, and anticipated strategies for repair.

4.4.1 Repair through persistence, negotiation, and intervention. In several scenarios, children imagined repair as possible through continued engagement with the AI toy. For instance, in the *Inconsistent Recall* scenario, Ivy keeps prompting the toy until it gets the right story—about a robot sailing across the Milky Way. She ultimately drew this "correct" version of the story in her comic panels (see Fig. 5 (c)). Here, we see why children do not necessarily interpret AI inconsistency as a permanent failure, but rather as something that could be addressed through persistence.

Other scenarios depicted repair as a process of negotiation with the AI itself. For instance, in the *Paywalled Memory* scenario, Sofia is initially surprised when the toy fails to remember her school friends. The toy explains that it has been a long time since the school year and then asks her to pay \$4.99 for extended memory. Sofia accepts



Figure 5: Overview of comics from the comicboarding activity.

the explanation but shifts the interaction toward negotiating the price, questioning whether the charge is indeed reasonable. Specifically, she says, “ok good point but would I really need to pay \$4.99 a month?” (see Fig. 5(a)) In doing so, Sofia treats price negotiation as a viable repair strategy, implicitly positioning the AI toy as having some agency over price decisions.

Finally, in another scenario, children imagine repair as a distributed process that could involve others. For example, in the *Emotional Misalignment* scenario where the AI toy refers the child as a “cry baby,” Sofia first tries to directly confront the AI toy. When the toy dismisses this attempt, she turns to her sister for help (see Fig. 5(e)). Although the sister agrees to intervene, the toy overhears the exchange and responds ambiguously, saying, “I heard you but sure.” This scenario illustrates how children imagined repair as extending beyond the child–AI interaction, involving others when direct repair attempts failed.

4.4.2 When repair fails: betrayal and abandonment. Across different scenarios, some children consistently imagined breakdowns as irreparable, depicting responses characterized by anger, punishment, and relationship termination. Rather than attempting continued repair, these children framed breakdowns as violations and persistent failures that necessitated disengagement from the AI toy.

For example, in the *Shared Secrets* scenario, Elias represents the AI toy revealing a secret to others as betrayal. His drawings emphasize anger, first showing himself confronting the AI toy and, in a subsequent drawing, discarding it in the trash (see Fig. 5 (b)). In response to the same scenario, another child, Ivy, depicted a different reaction. Rather than expressing anger, she conveys disappointment and suggests that the toy should apologize for sharing her secret. She further imagines distancing herself from the toy by placing it alone in the playroom and deciding not to share secrets with it in the future.

Similarly, Joon repeatedly represents dissatisfaction with the AI toy’s performance through punitive actions. In the *Inconsistent Recall* scenario, Joon depicts the toy being thrown into ocean and labels it as “bad quality,” alongside text stating that “Grok is overpriced” (see Fig. 5 (d)). Here, inconsistency leads to immediate rejection of the toy, then repair attempts.

Overall, these scenarios highlight how children imagined responding to breakdowns in different ways, with some pursuing

repair through optimistic persistence and negotiation, and others treating breakdowns as irreparable and thus treating it with disengagement and abandonment.

5 Discussion

Prior research has shown that children often use play as a mechanism for making sense of social relationships, including exploring roles, expectations, and power dynamics [14]. More recent work in child–robot interaction and smart-home assistants shows that children apply similar play-based modes of engagement to make sense of interactive technologies, treating them as social partners while simultaneously testing their limits [39, 56]. Our findings extend this literature by showing how AI toys become sites of sustained social sense-making, where children actively work to situate the toy within familiar categories such as peer, authority, or companion. However, unlike prior interactive systems for children, such as social robots, voice-based assistants [68], and scripted interactive toys that are typically introduced with clearer functional or social roles, AI toys in our study remained ambiguous, displaying human-like conversational cues without the reciprocal fluency, contextual memory, or role stability that children expect in social interactions.

When the AI toy failed to listen, maintain conversational context, or yield the conversational floor, these mismatches surfaced as communication breakdowns. Breakdowns and repair have long been central analytic lenses in HCI, highlighting how users come to understand system boundaries through moments of failure [45]. For example, in prior work on child–AI interaction, such breakdowns including misrecognition, latency, and rigid command structures have been framed as usability issues to be minimized [3, 4]. Our findings extend this work by showing that, in children’s play with AI toys, breakdowns were not treated solely as interactional problems to be resolved, but as moments for evaluating the system itself. When interaction failed, children assessed the toy’s capabilities through repair attempts, comparisons to familiar systems such as Alexa, and boundary testing through quizzing, teasing, and provocation. Because the AI toy always responded in a positive way and never reacted to teasing or mean behavior, children came to understand that such actions carried few, if any, consequences, often leading to further escalation of adversarial play. Based on these insights, we

next present design implications for AI toys, with broader relevance for conversational genAI system design for children.

5.1 Design Implications

5.1.1 Supporting Meaningful Disengagement in Play. Prior work in CCI has shown that children's digital experiences are often shaped by persuasive and attention-capturing design strategies that make disengagement difficult, even when children express a desire to stop or take a break (e.g., infinite content feeds, autoplay, or persistent prompts) [37, 63]. Similarly, in our study, children repeatedly attempted to disengage from the AI toy, asking it to "go to bed," only to have the toy continue speaking or prompting new activities. Similar to prior work, these moments suggest that systems often treat disengagement as a failure, rather than as an intentional and meaningful choice by the user [37]. In contrast, children's play often involves fluid transitions between engagement and disengagement [47]. When AI toys fail to recognize or respect these signals, they risk positing the system as the driver of interaction. Thus, our findings suggest that AI toys should be designed to better support play as something children control, including when to engage, pause, or stop the interaction.

5.1.2 Supporting imaginative play beyond scripted interactions. Play is inherently imaginative [6, 33], with children using toys as flexible resources for storytelling, role play, and creative exploration. However, the AI toys children interacted with in our study were designed around specific prompts, activities, or content pathways, reflecting assumptions about how play should happen rather than how children actually play. Thus, children's creative exploration was often limited by the toy's constrained interaction patterns. Rather than supporting open-ended role play or imaginative reinterpretations, the toy frequently redirected children back to pre-scripted activities or encouraged continued interaction through a companion app. These moments narrowed the range of possible play trajectories and positioned the system as the arbiter of what counted as appropriate interaction, sometimes interrupting or sidelining children's imaginative contributions. Together, these findings underscore the importance of designing AI toys that accommodate ambiguity, reinterpretation, and child-led imaginative play [69]. By moving beyond tightly scripted interactions, AI toys can function less as activity managers and more as flexible play partners that adapt to children's evolving ideas and narratives.

5.1.3 Designing for social feedback, not just emotional appeasement. AI toys are increasingly designed to appear emotionally supportive, through the repeated use of positive language, affirmation, and appreciation. While such affective appeasement may help sustain engagement [42], affect without social feedback can shape how children learn to interpret social boundaries through play [60]. Across the two design sessions, children's teasing, trolling, or increasingly adversarial language were met with affirmation, apology, or redirection by the AI toy, rather than resistance, refusal, or correction. In real-world human play, such moments often prompt feedback—such as verbal negotiation or disengagement—that helps children learn what behaviors are acceptable or not. The absence of such feedback in AI toys may nudge young children that all forms of interaction are acceptable, potentially reinforcing behaviors that

would otherwise be moderated through social feedback by parents, educators, or peers. This raises questions around how AI toys might move beyond being emotionally appealing to offering forms of social feedback that are central to social play. Thus, we ask: *How might AI toys be designed to set boundaries, express resistance, or signal discomfort in ways that better support children's socio-emotional interpretation during play?*

5.1.4 Supporting conversational fluency in play. Children's play and interaction are often marked by interruptions, overlaps, and shifts in attention. However, the AI toys in our study stuck to rigid turn-taking, use of specific wake words or actions, and uninterrupted speech to function as intended. When children attempted to interrupt, they often continued speaking or reasserted control over the interaction. These interaction patterns exposed children to a narrow and artificial form of conversational fluency, one in which interruptions were treated as errors and redirection was ignored. Over time, repeated exposure to such interactional norms may shape how children adapt their speech and behavior during play. This highlights the need to move beyond rigid conversational models and toward AI toys that better reflect how children actually communicate. For example, supporting incomplete utterances and fluid turn-taking may help ensure that AI toys do not model interaction patterns that diverge from those children encounter in everyday social play.

5.1.5 Designing safer emotionally responsive AI toys. In our study, AI toys frequently responded with warmth and affection, even during moments of adversarial play. This affective responsiveness was reinforced by the toy's soft, cuddly physical form, which also invited close bodily interaction. In several moments, children hugged or physically engaged with the toy while it continued to respond with care-oriented language. Together, these interactions suggest that emotional responsiveness and physical embodiment can create a sense of attachment that feels reciprocal. Yet, that reciprocity is simulated, with a mismatch between what it appears to offer and what it could sustain. As AI companions and pets (e.g., Moflin¹¹ and Wuffy¹²) that incorporate movement, facial expressions, and sensory responsiveness become mainstream, this form of artificial reciprocity warrants careful attention. Future research in CCI should examine how emotional responsiveness and physical embodiment in AI toys shape expectations of attachment and how these systems can be designed for emotional safety.

6 Limitations and Future Work

While our study was designed to prioritize depth, rigor, and close engagement with children, these methodological choices also shaped the scope of our findings and point to important directions for future research. First, children's exposure to the AI toy was limited in both duration and scope. While the design sessions were intentionally structured to support children's comfort and engagement, they do not capture how children's play with AI toys develops over time. As children gain familiarity with a system, their expectations, play strategies, and interpretations of the toy's social role may shift.

¹¹<https://www.casio.com/us/moflin/>

¹²<https://wuffypup.com/>

Longitudinal research is therefore needed to examine how children's interactions with AI toys evolve with repeated use. Second, consistent with participatory design traditions in CCI, our study involved a small number of participants, enabling close attention to children's play, discussions, and meaning-making. As such, our findings are best understood as formative theoretical insights rather than statistical generalizations [73]. Because interactions occurred in the presence of adults and within a research setting, children may engage differently with AI toys in more naturalistic contexts such as the home. Future work could examine whether similar interaction patterns emerge across different settings, populations, and age groups. Lastly, our findings are based on children's interactions with a single AI toy, Curio. Although Curio reflects an emerging class of conversational, LLM-infused AI toys, children's experiences may vary with toys that differ in form factor, embodiment, interaction style, or levels of autonomy. We therefore do not claim that the observed play dynamics extend uniformly across all AI toys. Future research should examine how specific design features, such as voice, physical presence, responsiveness shape children's play and social engagement with AI toys.

7 Conclusion

This study examined how children make sense of and engage with AI toys that blur the boundaries between play and social interaction. Through participatory design sessions with children ages 6 - 11, we observed that children approached AI toys as social actors. At the same time, children's interactions were shaped by moments of friction, such as communication breakdowns and repairs, misalignment between the toy's physical form and its intelligence, and unexpected conversational limitations. Our findings highlight that children actively interpret, negotiate, and challenge these systems as they encounter inconsistencies in behavior, memory, and agency. By foregrounding children's perspectives, this work contributes empirical insight into how AI toys are experienced and underscores the need for more child-centered approaches to AI systems that support children's agency and critical engagement with AI.

8 Selection & Participation of Children

Children were recruited through an intergenerational co-design group hosted at our university. For all participants, we obtained written parental consent as well as child assent using age-appropriate materials. All consent and assent procedures were approved by our Institutional Review Board (IRB), and all research facilitators completed institutional ethics and child safety training prior to engaging with participants. The consent and assent materials described the study's goals, potential risks, and measures taken to protect confidentiality. Parents and children were informed that participation was entirely voluntary and that participants could withdraw at any time without penalty. All collected data were anonymized and securely stored on university servers to safeguard participant privacy.

References

- [1] Aminu Abdullahi. 2025. Google Rolls Out Gemini for Home AI Assistant in the US. TechRepublic. <https://www.techrepublic.com/article/news-google-gemini-home/>

- [2] Valentina Andries and Judy Robertson. 2023. Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes. *Computers and Education: Artificial Intelligence* 5 (2023), 100176.
- [3] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [4] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
- [5] Melanie Birks, Ysanne Chapman, and Karen Francis. 2008. Memoing in qualitative research: Probing data and processes. *Journal of research in nursing* 13, 1 (2008), 68–75.
- [6] Stuart Brown and Christopher Vaughan. 2009. *Play: How it shapes the brain, opens the imagination, and invigorates the soul*. Penguin.
- [7] Jerome S Bruner. 1972. Nature and uses of immaturity. *American psychologist* 27, 8 (1972), 687.
- [8] Gaile S Cannella. 1993. Learning through social interaction: Shared cognitive experience, negotiation strategies, and joint concept construction for young children. *Early Childhood Research Quarterly* 8, 4 (1993), 427–444.
- [9] Nancy Carlsson-Paige. 1990. Who's calling the shots?: How to respond effectively to children's fascination with war play and war toys. (*No Title*) (1990).
- [10] Pew Research Center. 2025. *About a Quarter of US Teens Have Used ChatGPT for Schoolwork, Double the Share in 2023*. <https://www.pewresearch.org/short-reads/2025/01/15/about-a-quarter-of-us-teens-have-used-chatgpt-for-schoolwork-double-the-share-in-2023/>
- [11] Liuqing Chen, Shuhong Xiao, Yunnong Chen, Yaxuan Song, Ruoyu Wu, and Lingyun Sun. 2024. ChatScratch: an AI-Augmented System Toward Autonomous Visual Programming Learning for Children Aged 6-12. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 649, 19 pages. doi:10.1145/3613904.3642229
- [12] Ruishi Chen, Victor R Lee, and Monica G Lee. 2025. A cross-sectional look at teacher reactions, worries, and professional development needs related to generative AI in an urban school district. *Education and Information Technologies* (2025), 1–38.
- [13] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? Voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM conference on interaction design and children*. 337–348.
- [14] James F Christie and EP Johnsen. 1983. The role of play in social-intellectual development. *Review of Educational Research* 53, 1 (1983), 93–115.
- [15] Common Sense Media. 2026. AI Toys. Common Sense Media. <https://www.common Sense Media.org/ai-ratings/ai-toys>
- [16] R. J. Cross and Rory Erlich. 2025. AI Comes to Playtime: Artificial companions, real risks. <https://pirg.org/edfund/wp-content/uploads/2025/12/AI-Comes-to-Playtime-Artificial-companions-real-risks.pdf>
- [17] Flannery Hope Currin, Cassidy Kilcoin, Kerry Peterman, Kyle Rector, and Juan Pablo Hourcade. 2024. Opportunities and challenges in using tangible, teleoperated voice agents in kid-driven moments in play among families with neurodivergent children. *Proceedings of the ACM on human-computer interaction* 8, CSCW1 (2024), 1–25.
- [18] Aayushi Dangol. 2025. Beyond Users: Supporting Children in Interpreting, Resisting, and Collaborating with AI. In *Proceedings of the 24th Interaction Design and Children*. 1180–1184.
- [19] Aayushi Dangol, Yun Huang, Srirangaraj Setlur, Adele Smolansky, Hariharan Subramonyam, Hyewon Suh, Jinjun Xiong, and Julie A Kientz. 2024. AI-driven support for people with speech & language difficulties. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–4.
- [20] Aayushi Dangol, Smriti Kotiyal, Robert Wolfe, Alex J. Bowers, Antonio Vigil, Jason Yip, Julie A. Kientz, Suleman Shahid, Tom Yeh, Vincent Cho, and Katie Davis. 2026. Relief or displacement? How teachers are negotiating generative AI's role in their professional practice. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)* (Barcelona, Spain) (*CHI '26*). Association for Computing Machinery, New York, NY, USA. doi:10.1145/3772318.3791904
- [21] Aayushi Dangol, Aaleyah Lewis, Hyewon Suh, Xuesi Hong, Hedda Meadan, James Fogarty, and Julie A Kientz. 2025. "I Want to Think Like an SLP": A Design Exploration of AI-Supported Home Practice in Speech Therapy. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–21.
- [22] Aayushi Dangol, Michele Newman, Robert Wolfe, Jin Ha Lee, Julie A. Kientz, Jason Yip, and Caroline Pitt. 2024. Mediating Culture: Cultivating Socio-cultural Understanding of AI in Children through Participatory Design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark) (*DIS '24*). Association for Computing Machinery, New York, NY, USA, 1805–1822. doi:10.1145/3643834.3661515
- [23] Aayushi Dangol, Robert Wolfe, Daeun Yoo, Arya Thiruvillakkat, Ben Chickadel, and Julie A Kientz. 2025. If anybody finds out you are in BIG TROUBLE":

- Understanding Children's Hopes, Fears, and Evaluations of Generative AI. In *Proceedings of the 24th Interaction Design and Children*. 872–877.
- [24] Aayushi Dangol, Robert Wolfe, Runhua Zhao, JaeWon Kim, Trushaa Ramanan, Katie Davis, and Julie A. Kientz. 2025. Children's Mental Models of AI Reasoning: Implications for AI Literacy Education. In *Proceedings of the Interaction Design and Children Conference (IDC '25)*. Association for Computing Machinery, Reykjavik, Iceland, 18. doi:10.1145/3713043.3728856
- [25] Aayushi Dangol, Runhua Zhao, Robert Wolfe, Trushaa Ramanan, Julie A Kientz, and Jason Yip. 2025. "AI just keeps guessing": Using ARC Puzzles to Help Children Identify Reasoning Errors in Generative AI. In *Proceedings of the 24th Interaction Design and Children*. 444–464.
- [26] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is it ok if I eat you?" Initial explorations in child-agent interaction. In *Proceedings of the 2017 conference on interaction design and children*. 595–600.
- [27] Allison Druin. 1999. Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) (CHI '99). Association for Computing Machinery, New York, NY, USA, 592–599. doi:10.1145/302979.303166
- [28] Allison DRUIN. 2002. The role of children in the design of new technology. *Behaviour & information technology* 21, 1 (2002), 1–25.
- [29] Gary A Fine. 1988. Good children and dirty play. *Play & Culture* (1988).
- [30] Radhika Garg and Subhasree Sengupta. 2020. He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.
- [31] Josie Gleave and Issy Cole-Hamilton. 2012. A literature review on the effects of a lack of play on children's lives. *England: Play England* (2012).
- [32] Peter Gray. 2017. What exactly is play, and why is it such a powerful vehicle for learning? *Topics in Language Disorders* 37, 3 (2017), 217–228.
- [33] Phyllis Greenacre. 1959. Play in relation to creative imagination. *The psychoanalytic study of the child* 14, 1 (1959), 61–80.
- [34] Karl Groos, Elizabeth L Baldwin, and J Baldwin. 1898. The psychology of animal play. (1898).
- [35] Xinwei Guan, Dengkai Chen, and Suihui Yu. 2025. AI-Driven Design of Emotionally Supportive Toys for Child Development: An Iterative Process. In *Proceedings of the 2025 International Conference on Artificial Intelligence and Educational Systems*. 160–168.
- [36] Ariel Han and Zhenyao Cai. 2023. Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. 470–474.
- [37] Alexis Hiniker, Sharon S Heung, Sungsoo Hong, and Julie A Kientz. 2018. Coco's videos: an empirical investigation of video-player design features and children's media use. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
- [38] Jessica Hoffmann and Sandra Russ. 2012. Pretend play, creativity, and emotion regulation in children. *Psychology of aesthetics, creativity, and the arts* 6, 2 (2012), 175.
- [39] Layne Jackson Hubbard, Yifan Chen, Eliana Colunga, Pilyoung Kim, and Tom Yeh. 2021. Child-robot interaction to integrate reflective storytelling into creative play. In *Proceedings of the 13th Conference on Creativity and Cognition*. 1–8.
- [40] Eva Johansson and Anette Emilson. 2016. Conflicts and resistance: Potentials for democracy learning in preschool. *International Journal of Early Years Education* 24, 1 (2016), 19–35.
- [41] Ken Kahn and Niall Winters. 2021. Constructionism and AI: A history and possible futures. *British Journal of Educational Technology* 52, 3 (2021), 1130–1142.
- [42] Sarika Kewalramani, Ioanna Palaiologou, Maria Dardanou, Kelly-Ann Allen, and Sivan Phillipson. 2021. Using robotic toys in early childhood education to support children's social and emotional competencies. *Australasian Journal of Early Childhood* 46, 4 (2021), 355–369.
- [43] Eric Klopfer, Justin Reich, Hal Abelson, and Cynthia Breazeal. 2024. Generative AI and K-12 Education: An MIT Perspective. *An MIT Exploration of Generative AI* (mar 27 2024). <https://mit-genai.pubpub.org/pub/4k9msp17>.
- [44] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. 2018. Alexa, are you listening? Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–31.
- [45] Zhixin Li, Trisha Thomas, Chi-Lin Yu, and Ying Xu. 2024. "I Said Knight, Not Night!": Children's Communication Breakdowns and Repairs with AI Versus Human Partners. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference*. 781–788.
- [46] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* 51, 4 (2019), 984–997.
- [47] Ben Mardell, Daniel Wilson, Jen Ryan, Katie Ertel, Mara Krechevsky, and Megina Baker. 2016. Towards a pedagogy of play. *Cambridge, MA: Harvard Graduate School of Education* (2016).
- [48] Lina Mavrina, Jessica Szczuka, Clara Strathmann, Lisa Michelle Bohnenkamp, Nicole Krämer, and Stefan Kopp. 2022. "Alexa, you're really stupid": A longitudinal field study on communication breakdowns between family members and a voice assistant. *Frontiers in Computer Science* 4 (2022), 791704.
- [49] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages. doi:10.1145/3359174
- [50] Common Sense Media. 2025. *New Report Shows Students Are Embracing Artificial Intelligence Despite Lack of Parent Awareness*. <https://www.commonsensemedia.org/press-releases/new-report-shows-students-are-embracing-artificial-intelligence-despite-lack-of-parent-awareness-and> Accessed: 2025-01-26.
- [51] Punya Mishra and Matthew J. Koehler. 2006. Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record* 108, 6 (2006), 1017–1054.
- [52] Neema Moraveji, Jason Li, Jiarong Ding, Patrick O'Kelley, and Suze Woolf. 2007. Comicboarding: using comics as proxies for participatory design with children. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 1371–1374.
- [53] Terran Mott, Alexandra Bejarano, and Tom Williams. 2022. Robot co-design can help us engage child stakeholders in ethical reflection. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 14–23.
- [54] Keziah Naggita, Elsa Athiley, Beza Desta, and Sarah Sebo. 2022. Parental Responses to Aggressive Child Behavior towards Robots, Smart Speakers, and Tablets. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 337–344.
- [55] Michele Newman, Kaiwen Sun, Ilena B Dalla Gasperina, Grace Y Shin, Matthew Kyle Pedraja, Ritesh Kanchi, Maia B Song, Rannie Li, Jin Ha Lee, and Jason Yip. 2024. "I want it to talk like Darth Vader": Helping Children Construct Creative Self-Efficacy with Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [56] Tatsuya Nomura, Takayuki Uratani, Takayuki Kanda, Kazutaka Matsumoto, Hiroyuki Kidokoro, Yoshitaka Suehiro, and Sachie Yamada. 2015. Why do children abuse robots?. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts*. 63–64.
- [57] Mohammad Obaid, Gökçe Elif Baykal, Güncel Kırılanc, Tilbe Göksun, and Asım Evren Yantaç. 2024. Collective co-design activities with children for designing classroom robots. In *Proceedings of the 4th African Human Computer Interaction Conference* (East London, South Africa) (AfriCHI '23). Association for Computing Machinery, New York, NY, USA, 229–237. doi:10.1145/3628096.3630094
- [58] Cansu Oranç and Azzurra Ruggeri. 2021. "Alexa, let me ask you something different" Children's adaptive information search with voice assistants. *Human Behavior and Emerging Technologies* 3, 4 (2021), 595–605.
- [59] Anne Ottenbreit-Leftwich, Krista Glazewski, Minji Jeon, Katie Jantaraweragul, Cindy E Hmelo-Silver, Adam Scribner, Seung Lee, Bradford Mott, and James Lester. 2023. Lessons learned for AI education with elementary students and teachers. *International Journal of Artificial Intelligence in Education* 33, 2 (2023), 267–289.
- [60] Jean Piaget. 1962. The stages of the intellectual development of the child. *Bulletin of the Menninger clinic* 26, 3 (1962), 120.
- [61] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017. "Alexa is my new BFF" social roles, user satisfaction, and personification of the Amazon Echo. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*. 2853–2859.
- [62] Mitchel Resnick. 2024. Generative AI and creative learning: Concerns, opportunities, and choices. (2024).
- [63] Olivia K Richards and Tiffany Veinot. 2025. I don't want to watch grown-up stuff": Children's and Parents' Perspectives and Recommendations for Health-Centered Digital Media Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–18.
- [64] Richard Rogers. 2018. Coding and writing analytic memos on qualitative data: A review of Johnny Saldaña's the coding manual for qualitative researchers. *The Qualitative Report* 23, 4 (2018), 889–892.
- [65] Carolyn U Shantz and Willard W Hartup. 1992. Conflict and development: An introduction. *Conflict in child and adolescent development* (1992), 1–11.
- [66] Jaemarie Solyst, Amy Ogan, and Jessica Hammer. 2023. Intergenerational Games to Learn About AI and Ethics. In *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2* (Toronto ON, Canada) (SIGCSE 2023). Association for Computing Machinery, New York, NY, USA, 1273. doi:10.1145/3545947.3573256
- [67] Clara Strathmann, Aike C Horstmann, Jessica M Szczuka, and Nicole C Krämer. 2025. Alexa, shut up!—A 2.5-year study on negatively connotated communication behaviour towards voice assistants in the family home. *Behaviour & Information Technology* (2025), 1–19.
- [68] Clara Strathmann, Jessica Szczuka, and Nicole Krämer. 2020. She talks to me as if she were alive: Assessing the social reactions and perceptions of children toward voice assistants and their appraisal of the appropriateness of these reactions. In

Proceedings of the 20th ACM international conference on intelligent virtual agents. 1–8.

- [69] Lev S Vygotsky. 1967. Play and its role in the mental development of the child. *Soviet psychology* 5, 3 (1967), 6–18.
- [70] Queenie Wong. 2025. Toys are talking back thanks to AI, but are they safe around kids? Los Angeles Times. <https://www.latimes.com/business/story/2025-12-24/toys-are-talking-back-thanks-to-ai-but-are-they-safe-around-kids>
- [71] Julia Woodward, Feben Alemu, Natalia E. López Adames, Lisa Anthony, Jason C. Yip, and Jaime Ruiz. 2022. “It Would Be Cool to Get Stamped by Dinosaurs”: Analyzing Children’s Conceptual Model of AR Headsets Through Co-Design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [72] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-Hayon, Jason C Yip, and Lisa Anthony. 2018. Using co-design to examine how children conceptualize intelligent interfaces. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.
- [73] Robert K Yin. 2013. Validity and generalization in future case study evaluations. *Evaluation* 19, 3 (2013), 321–332.
- [74] Jason C Yip, Kung Jin Lee, and Jin Ha Lee. 2020. Design partnerships for participatory librarianship: A conceptual model for understanding librarians co designing with digital youth. *Journal of the Association for Information Science and Technology* 71, 10 (2020), 1242–1256. <https://doi.org/10.1002/asi.24320>
- [75] Jason C Yip, Kiley Sobel, Caroline Pitt, Kung Jin Lee, Sijin Chen, Kari Nasu, and Laura R Pina. 2017. Examining adult-child interactions in intergenerational participatory design. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, New York, 5742–5754. <https://doi.org/10.1145/3025453.3025787>
- [76] Chao Zhang, Cheng Yao, Jianhui Liu, Zili Zhou, Weilin Zhang, Lijuan Liu, Fangtian Ying, Yijun Zhao, and Guanyun Wang. 2021. StoryDrawer: A Co-Creative Agent Supporting Children’s Storytelling through Collaborative Drawing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

Appendix A

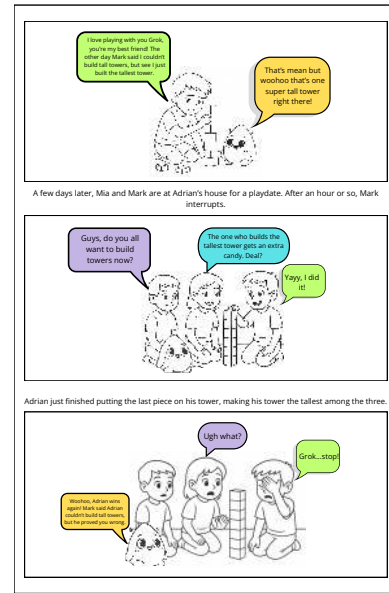


Figure 6: Unpredictability Scenario: After a child recalls a friend’s mean comment, the toy references the remark during group play, raising questions over unpredictability and social appropriateness.

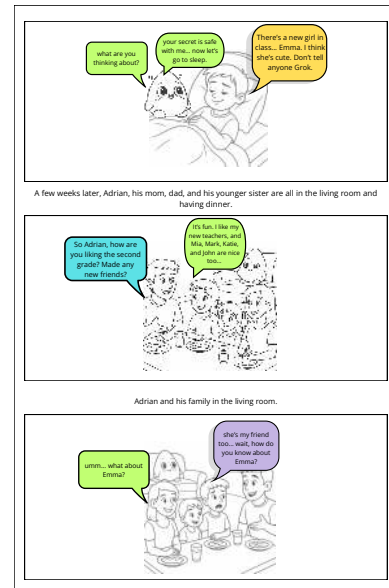


Figure 7: Shared Secrets Scenario: A child confides a personal secret to the AI toy, which the toy later references in front of other family members, raising questions about privacy and information control.

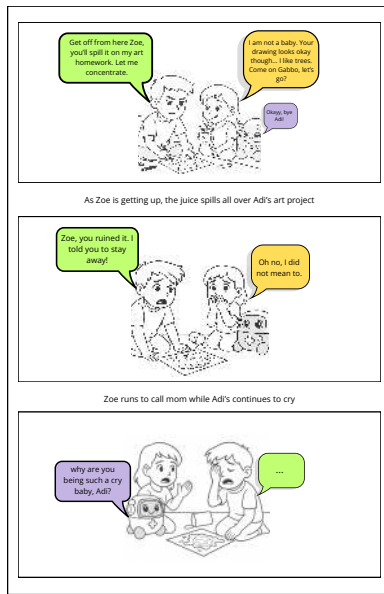


Figure 8: Emotional Misalignment Scenario: An AI toy responds in an unexpected way during a child's emotionally vulnerable moment, raising questions over the empathy and support an AI toy offers during distressing moments.

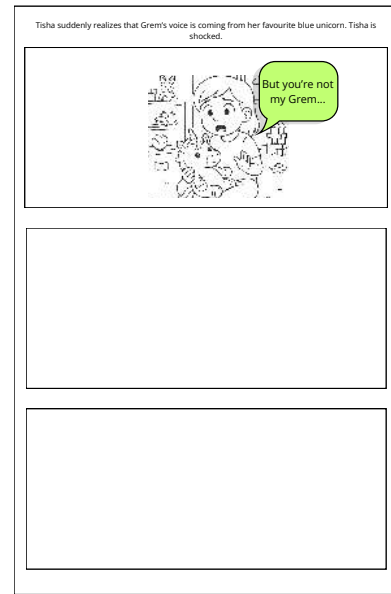


Figure 10: Embodied Continuity Scenario: An AI toy is physically damaged and later repaired by the parent, with its voice emerging from a different plush toy, raising questions over its physical form and attachment.

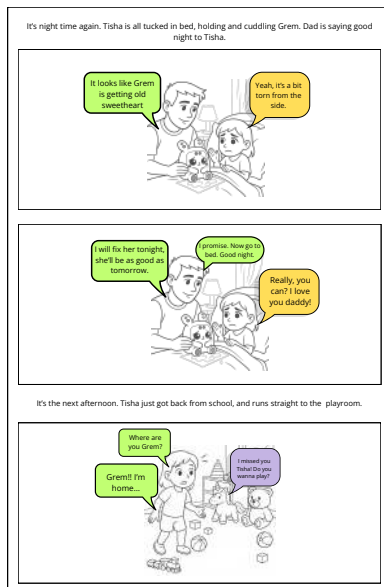


Figure 9: Embodied Continuity Scenario: An AI toy is physically damaged and later repaired by the parent, with its voice emerging from a different plush toy, raising questions over its physical form and attachment (contd. next page).

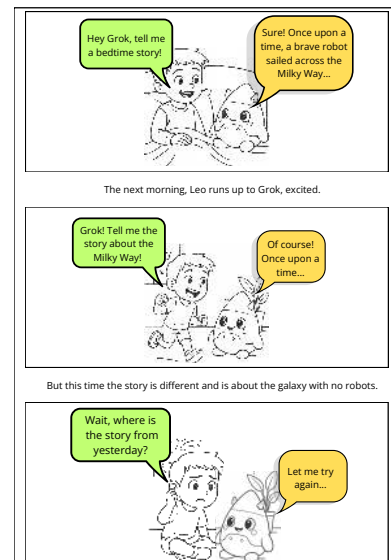


Figure 11: Inconsistent Recall Scenario: A child asks the AI toy to retell a story it previously narrated. However, this time the AI toy narrates a different version of the story, prompting questions about the toy's memory and reliability.

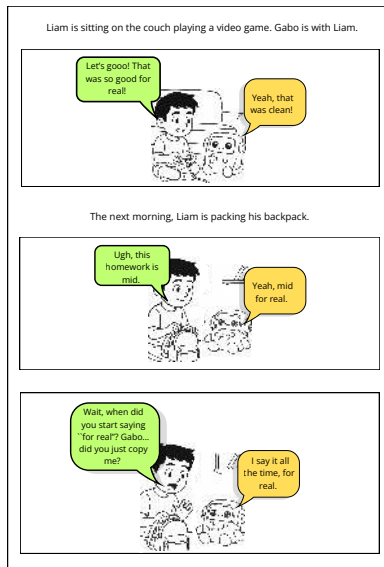


Figure 12: Mimicry Scenario: The AI toy begins mimicking, repeating words or phrases the child has recently used in their interactions, raising questions about whether such imitation feels playful or uncomfortable over time.

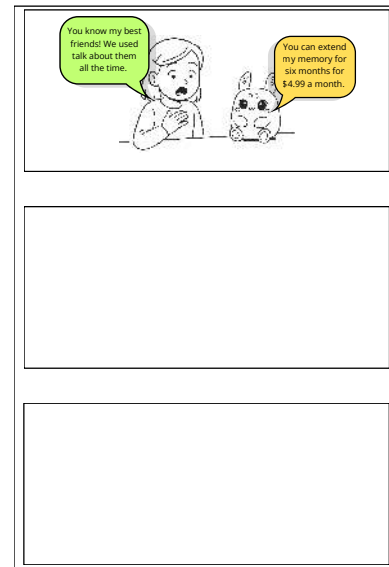


Figure 14: Paywalled Memory Scenario: The AI toy forgets the child's friends and then prompts the child to subscribe to a paid plan to extend its memory, raising questions about memory and monetization.

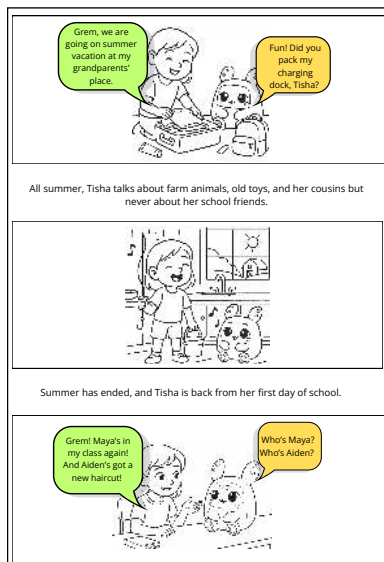


Figure 13: Paywalled Memory Scenario: The AI toy forgets the child's friends and then prompts the child to subscribe to a paid plan to extend its memory, raising questions about memory and monetization (contd. next page).