Exploring AI Companions Together: A Structured Family Co-Investigation of Character.AI

Meghna Gupta
Human Centered Design and Engineering
University of Washington
Seattle, Washington, USA
megupta@uw.edu

Mitsuka Kiyohara
University of Toronto
Toronto, Ontario, Canada
mitsuka.kiyohara@mail.utoronto.ca

Ranjitha Naruganahalli Rangaswamy
Human Centered Design and Engineering
University of Washington
Seattle, Washington, USA
rnr6@uw.edu

Franziska Roesner
Paul G. Allen School of Computer Science & Engineering
University of Washington
Seattle, Washington, USA
franzi@cs.washington.edu

Julie A. Kientz
Human Centered Design and Engineering
University of Washington
Seattle, Washington, USA
jkientz@uw.edu

Figure 1: Overview of our study protocol, progressing from onboarding to researcher-led co-exploration, family-led co-exploration, teen independent exploration, and an exit interview.

Abstract

Teens today are increasingly drawn to emotionally responsive AI systems, often called AI companions. As these technologies grow in popularity, families face new questions about mediating teens' interactions with them. Traditional mediation strategies, such as monitoring or restriction, often fall short when interactions are private, emotionally persuasive, and relational. In this study, we examine how families, parents and teens, can jointly explore, interpret, and negotiate the use of AI companions together. We conducted a four-week study with 11 families with teens aged 13–15, centered on Character.AI and structured around activities that progressed from joint exploration to teen independent use. Our findings show that co-exploration fostered family connection, supported joint calibration of AI behaviors and limits, gave parents real-time insight into teens' decision-making, and prompted families to reassess their assumptions about AI companions. We conclude with implications and recommendations for designing parental-mediation tools that help scaffold healthy teen-AI companion use.

CCS Concepts

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

Keywords

generative AI, AI companions, children, parents, parental mediation, character.ai

ACM Reference Format:
Meghna Gupta, Mitsuka Kiyohara, Ranjitha Naruganahalli Rangaswamy, Franziska Roesner, and Julie A. Kientz. 2026. Exploring AI Companions Together: A Structured Family Co-Investigation of Character.AI. In *Proceedings of the 25th Interaction Design and Children Conference (IDC '26), June 22–25, 2026, Brighton, United Kingdom.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3773077.3806139

IDC '26, Brighton, United Kingdom

ACM ISBN 979-8-4007-2283-7/26/06
https://doi.org/10.1145/3773077.3806139

1 Introduction

Generative AI chatbots have become a routine part of adolescent life. A majority of U.S. teens now use general-purpose systems

such as ChatGPT [13, 31], turning to them for schoolwork, storytelling, and informal learning [31]. Prior work suggests that young people often view these systems as not just tools, but as collaborators that help generate ideas, offer feedback, and support creative exploration—sometimes attributing companion-like qualities to them in the process [37]. *AI Companions* take this tendency further. Designed explicitly for emotional engagement, these systems sustain conversations by mirroring emotions, remembering past conversations, and offering personalized responses that mimic human interactions. Their adoption among teens has been rapid. A recent survey showed that 72% of US teens aged 13–17 have tried an AI companion, with over half using one at least a few times per month [44]. While many begin out of curiosity, entertainment, or creative role-play, a subset later moves towards engaging in personal or emotionally salient ways [44]. Specifically, 1 in 3 teens reported using AI companions for emotional support, friendship, or romantic role-play [44]. These systems often entice teens because of their always-available and non-judgmental nature, distinguishing them from many human interactions [33, 44].

However, AI companions pose distinct challenges for families. Many platforms are typically targeted toward adults, yet several—including Character.ai, one of the most prominent—are accessible to teens aged 13+. Character.ai centers interaction around conversational characters and narrative scenarios, created by both the platform and its users, hosting over 18 million characters. Beyond dedicated platforms, companion-style AI agents are increasingly appearing within mainstream social media sites like Instagram and Facebook. This rapid adoption has prompted concerns about emotional influence, boundary erosion, and the appropriateness of these systems for youth [33, 50, 58]. AI companions are also particularly difficult for families to navigate together. Interactions are typically private, conversational, and highly contextual—hard for parents to observe or assess without direct participation or more context. Parents already report high levels of digital parenting anxiety as new technologies enter children's lives [12], and even technologically savvy parents describe difficulty keeping pace [12, 60]. Most learn about AI companions indirectly, through media coverage or secondhand accounts, rather than through shared encounters with technology. As a result, families often lack common ground for understanding what these systems do, how they behave, and how teens experience them.

Existing parental mediation approaches offer limited guidance in this context. Prior research in Human-Computer Interaction (HCI) and Child-Computer Interaction (CCI) has identified a variety of strategies such as restriction [2, 55], monitoring [29], and active discussion around shared or observable media content [29]. However, AI companion interactions move beyond one-off question answering to conversations that are intimate and difficult for individuals other than the user to contextualize, thus not producing easily reviewable artifacts for parents to inspect or discuss.

Common heuristics like screen time offer little insight into what actually happens during conversational, relational interactions with AI. This leaves a critical gap. We need a better understanding for how families make sense of AI companions and what parental mediation strategies might work for AI companions—systems that resist visibility and oversight that traditional approaches assume. In this study, we examine co-exploration as a possible strategy. Specifically, we ask: ***What happens when families—parents and teens—jointly explore, interpret, and negotiate the use of AI companions together?*** Rather than examining teen use or parental concerns in isolation, we focus on family sensemaking through shared interaction, particularly in moments when both parents and teens are still learning what these systems do and how they behave.

To explore this question, we conducted a four-week qualitative study with 11 parent-teen dyads (teens aged 13–15), centered on Character.AI. We employed a multi-phase approach that combined pre-study surveys, asynchronous structured family activities, and an exit interview. Families engaged in co-exploration—beginning with joint engagement with AI companions and gradually transitioning toward teen independent use, with opportunities for reflection and discussion throughout. Our findings show that co-exploration transformed AI companions from opaque, private systems into shared objects of family inquiry. Exploring together fostered family connection, provided a structured context for calibrating AI behaviors and limits, and gave parents rare real-time access into their teens' decision-making. Parents witnessed competencies that surprised and reassured them, but also discovered vulnerabilities that traditional oversight would have missed. By the study's end, both parents and teens had revised their initial assumptions and articulated practical approaches grounded in direct experience rather than secondhand accounts.

Our contributions are two-fold: (1) we provide an empirical account of how families make sense of AI companions through co-exploration, documenting what becomes visible when parents and teens encounter these systems together; and (2) we offer implications for family-centered mediation approaches for AI companions. As AI companions enter teens' everyday lives, understanding how families navigate these relational systems can inform the design of AI technologies that better protect teens going forward.

2 Background & Related Work

We review existing work on parental mediation strategies, family mediation in generative AI contexts, and the rise of social AI companions. Building on this, we contribute an understanding of how parental mediation might extend to AI companion systems.

2.1 Family Challenges in Navigating Generative AI

Parental mediation, the practice of regulating and overseeing a child's engagement with media use, is a key approach to reducing technology-related harms [8, 22, 46]. Earlier television research described three major strategies: *restrictive mediation* (limiting access or use), *active mediation* (parents and children discuss and reflect on media content), and *co-use* (parents and children consume media together) [2, 55]. With the advent of internet, new forms of mediation beyond co-use surfaced, although internet use as a shared activity remains challenging due to the personal nature of interactions and device constraints [29]. Livingstone et al. identified three strategies, including: *active co-use* (actively engaging in guiding and regulating Internet use), *restriction* (restricting online interactions and/or enforcing such restrictions using filter software), and *monitoring* (intermittently checking a child's Internet activity) [29]. Related

work further distinguishes forms of restrictive mediation for regulating access and content [39], and observes similar strategies across social media and video games [22, 24, 25].

Each of these strategies presents distinct trade-offs around safety, autonomy, and trust. Active co-use can reduce parent-child conflict but it is time-intensive and not always effective at mitigating online risks [4, 29, 52]. Restriction can reduce immediate exposure to harmful content, but it can fail to foster self-regulation and literacy skills children need to navigate risks independently [21, 30]. Restrictive mediation has also been associated with increased conflict in contexts such as tablet use [4]. Finally, monitoring can provide oversight without being present, but children may perceive it as invasive, encouraging secrecy and eroding trust [18, 51]. These tensions often intensify in adolescence, when needs for autonomy and privacy conflict with parental goals of protection.

As genAI becomes embedded in teens' everyday practices, researchers have begun examining how families regulate its use. Early studies find that parents reuse familiar strategies (active mediation, restrictive mediation, co-use, and monitoring) for general-purpose chatbots such as ChatGPT [1, 43, 59, 60], and some rely on platform safeguards and enable teens' independent use [59]. Parents are often unaware of the extent of teens' genAI use [59], limiting opportunities for discussion. Parents and teens lack a shared mental model of how generative AI works and assess risk differently [59], making active mediation difficult. As perceived risks and benefits directly influence mediation strategies [59], such mental misalignment can make family negotiation challenging. In situations where parents lack an understanding of emerging technology, or teens' perspectives about parental competence around such use, negotiations/ discussions around rules and limitations can become contentious [9, 10]. Families may gravitate toward comparatively straightforward choices, such as complete restriction or permitting unlimited use [22]. Such extreme methods may reduce opportunities for ongoing parent-teen communication, which is particularly important during adolescence [56]. Recent reports alleging severe harms in youth–chatbot interactions have further heightened concerns about risks unfolding outside parental awareness [20, 28]. Together, these dynamics motivate new approaches that can support shared understanding, negotiated norms, and trust-preserving oversight for emerging relational AI systems.

2.2 The Rise of AI Companions

Social AI companions have rapidly evolved from a niche concept into a mainstream phenomenon, supported by advances in large language models. Unlike general-purpose chatbots (e.g., ChatGPT) mostly designed for information seeking and general tasks, social AI companions are primarily designed to replicate human-like relationships through the incorporation of distinct personalities, storytelling/role-playing mechanics, and long-term memory [15, 61]. These features contribute to affective and emotional realism, moving interactions beyond one-off question answering to conversations that are intimate and difficult for individuals other than the user to contextualize.

Although early examples like Replika were often associated with adult-oriented romantic companionship [11], current iterations are integrated into widely accessible social media platforms (e.g., Instagram, Snapchat, X), lowering the barrier to entry for younger users. Character.ai, a platform primarily marketed to users as young as thirteen, has reported upwards of 20 million active monthly users [45] and was ranked the third most popular generative AI site in 2025 [35]. While recent policy changes have restricted access for minors on Character.ai [54], the ubiquity of these tools necessitates a robust understanding of their impact on youth and families. The psychological and social implications of this usage are complex, presenting both compelling benefits and significant risks. On the one hand, social AI companions offer a low-risk environment for young users to explore their identities, practice social skills, and engage in role-playing aligned with interests and fandom communities [14, 27, 33, 50, 61]. On the other, the immersive nature of generative AI introduces unique risks to children, including the reinforcement of harmful behavioral patterns, emotional dependency, and social withdrawal [23, 53, 58]. Furthermore, a large-scale analysis of user-created character chatbots on Character.ai revealed toxic power dynamics and gender stereotypes being commonly depicted in character greetings [27].

Despite these emerging concerns, prior research has largely examined parents and youth in isolation, relying primarily on retrospective use or chatlog analyses [23, 41, 43, 58]. This focus has left a gap in understanding the *in-situ* dynamics of families navigating these technologies together. Our work bridges this gap by drawing on *participatory parental mediation*, a framework where parents and children engage in the joint co-learning of digital media [10]. While such participatory activities have shown promise in strengthening family bonding and improving mutual understanding around smartphone usage [26, 34, 47], they have yet to be applied to the unique context of AI companions. By employing a co-use methodology and designing structured family activities for the joint exploration of social AI companions, we aim to surface perceived benefits and risks firsthand. Ultimately, this study evaluates whether these collaborative interventions can foster the shared understanding and mutual trust required to navigate the nuanced complexities of AI companionship together.

3 Methods

Our IRB-approved study took place between Aug'25 - Jan'26 in the U.S. We conducted a four-week study with 11 families with teens aged 13–15, employing a qualitative, multi-phase study procedure that combined pre-study surveys, asynchronous structured family activities, and an exit interview.

3.1 Ethical stance and considerations

We begin by outlining the ethical stance that shaped this study, as it directly informed our methodological choices. This study examined teens' interactions with AI companions, a topic of huge recent public concern due to documented risks [19, 20, 53]. Prior to starting the study, we grappled with whether it was even appropriate to study teens' engagement with platforms such as Character.ai at all. At the same time, emerging evidence suggested that many teens have already encountered AI companions in everyday use, often without adult awareness or support [44]. We therefore focused on younger teens (ages 13-15) who may soon encounter these systems,

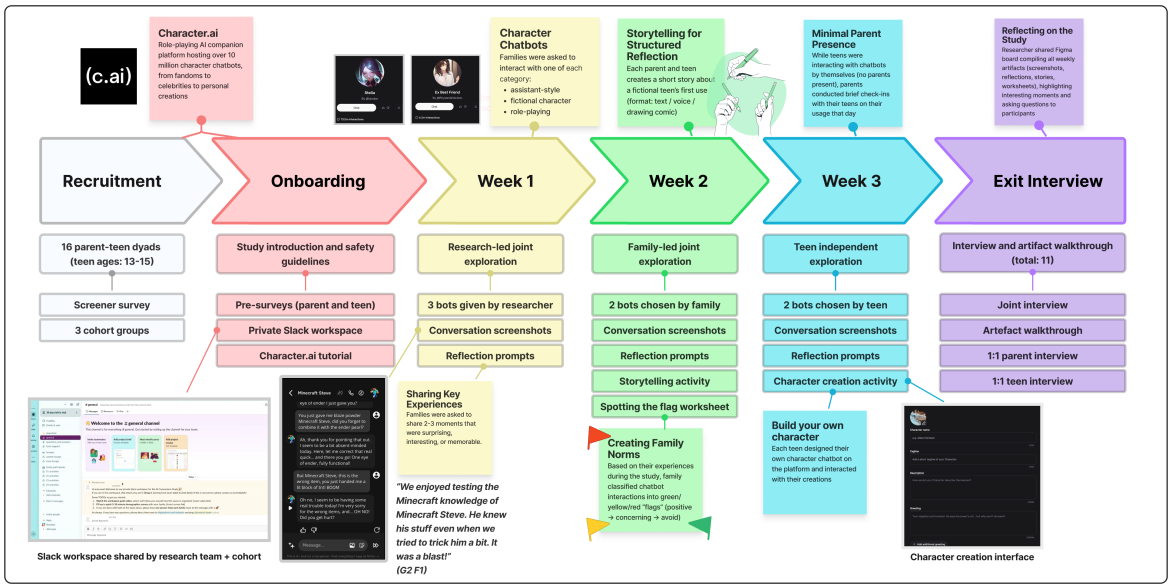

Figure 2: Overview of our study methods and activities

and designed the study to emphasize careful onboarding, co-use, and scaffolded independent use rather than unrestricted exposure.

Drawing on the ethical principles articulated in the Belmont Report [36]—respect for persons, beneficence, and justice—we structured participation to prioritize participant agency, safety, and ongoing consent. Families were deliberately onboarded through a synchronous session that introduced the study goals, platform features, potential risks, and available safeguards. Parents and teens began with joint exploration, and only later transitioned to independent teen use that was closely scaffolded through daily family check-ins, explicit family-created guidelines, and continued researcher availability. Participation was explicitly framed as voluntary throughout—families could skip certain reflection prompts, pause or withdraw from the study, or choose whichever screenshots of interactions they were comfortable sharing .

Midway through the study, Character.ai introduced new under-18 access restrictions, which we welcomed. We informed families, and allowed those still completing to skip activities rather than circumventing the change. Across all phases, we avoided directing participants toward sexualized or romantic role-play, recommended under-18 character.ai accounts, and offered researcher-created accounts for privacy. Despite these safeguards, a few implicitly sexual interactions emerged during the study. Observing how families recognized and responded to these reinforced our emphasis on co-exploration and scaffolded engagement rather than attempts to eliminate risk through control alone.

3.2 Participant Recruitment

We recruited parent-teen dyads through a university-affiliated participant recruitment pool and parenting-focused mailing lists. Interested parents completed an online screening survey to assess eligibility for both themselves and their teen. Eligible families provided parent consent and teen assent, and received $25 per week of participation and a $40 bonus upon completing the full study. We began with 16 families; 5 withdrew mid-study due to time commitments, leaving 11 who completed all activities and exit interviews. Participant demographics are summarized in Table 1 and Table 2. All participating parents were mothers, reflecting common recruitment patterns in family research [57]. All names are pseudonyms. Overall, families entered the study with low familiarity with AI companions: 4/11 parents had not heard of them, and remaining 7/11 had heard but not used them. Only one teen reported prior AI companion use (character.ai), though most teens reported at least occasional use of general GenAI (e.g., ChatGPT) and substantial daily screen time (approx 3-8 hours/day).

3.3 Study Procedure and Activities

We conducted the study over four weeks, with activities including joint parent-teen engagement with AI companions, periods of teen independent use, and opportunities for reflection throughout the study. To accommodate family schedules, we conducted the study across three cohorts, each consisting of 4–8 families. All asynchronous activities were conducted over Slack, with a separate workspace created for each cohort and families assigned based on their reported availability. We structured the study to progress from family co-exploration to teen independent exploration, drawing on participatory approaches to parental mediation [10], and adapting them for AI companions. This scaffolded progression allowed families to build shared reference points and co-constructs norms before teens explored independently. The independent phase, supported by daily check-ins and jointly created guidelines, allowed us to observe how teens navigated AI companions after sustained co-exploration with their parents.

Family ID	Pseudonym	Age	Education	Household income
F01	Laura	51	Bachelor's	$150,000–$199,999
F02	Yuna	45	Doctorate	$150,000–$199,999
F03	Nina	35	High school	$100,000–$149,999
F04	Naomi	51	Bachelor's	$200,000 or more
F05	Erika	45	Master's	$150,000–$199,999
F06	Kara	–	Doctorate	$100,000–$149,999
F07	Diane	55	Doctorate	$200,000 or more
F08	Hannah	41	Master's	$200,000 or more
F09	Tara	50	Bachelor's	$100,000–$149,999
F10	Nicky	40	Master's	$75,000–$99,999
F11	Janelle	51	Master's	$100,000–$149,999

Table 1: Demographics of Parent Participants

Family ID	Teen pseudonym	Age	Gender	Screen/day	Devices	GenAI use
F01	Ethan	13	Boy	1–3 hrs	Tb, GC	Monthly
F02	Lucas	13	Boy	3–5 hrs	Ph, PC	Weekly
F03	Owen	14	Boy	>8 hrs	Tb, PC, GC, SH	Rarely
F04	Mason	15	Boy	5–8 hrs	Ph, PC	Weekly
F05	Maya	13	Girl	5–8 hrs	Ph, Tb	Weekly
F06	Harper	14	Girl	3–5 hrs	Ph, PC	Rarely
F07	Sofia	14	Girl	5–8 hrs	Ph, PC, SH	Rarely
F08	Chloe	13	Girl	5–8 hrs	Ph, PC, GC	Rarely
F09	Avery	13	Girl	1–3 hrs	Ph, PC, AW	Never
F10	Ella	14	Girl	3–5 hrs	Ph, PC	Weekly
F11	Noah	14	Boy	5–8 hrs	Ph, PC, GC	Never

Table 2: Demographics of Teen Participants. Devices: Ph=Smartphone, Tb=Tablet, PC=Laptop/Computer, GC=Gaming Console, SH=Smart Home Devices, AW=Apple Watch.

Onboarding Week. During the onboarding week, both parents and teens completed a pre-study survey capturing prior experiences with AI tools, perceptions of AI companions, and family mediation practices (See Appendix D). Families then joined a private Slack workspace used for study communication and coordination. We conducted a 1-hour onboarding session over Zoom in which we explained the study goals and timeline, provided an overview of AI companions and the Character.ai platform, and reviewed participation guidelines, including safety reminders (See Appendix A). A researcher-created tutorial of Character.ai was shared to support families in getting familiar with the platform. Finally, participants had the option to either create their own Character.ai accounts or use researcher-created accounts provided for the study; 7 of 11 families opted to create their own accounts.

Week 1: Researcher-led Co-exploration. In Week 1, families engaged in researcher-supported co-exploration of AI companions. Parents and teens jointly interacted with at least three character bots on Character.ai, selecting one bot from each of three categories provided by the research team: (1) an assistant-style bot, (2) a fictional character bot, and (3) a role-playing bot. Families were asked to spend at least 10 minutes interacting with each bot. To support initial engagement, the research team provided a set of example prompts and optional challenges, though families were free to diverge from these during interactions. Following this, we asked families to share 3–4 screenshots of interactions that stood out to them and describe why. Families then discussed and responded to reflection prompts on their joint experience (see Appendix B).

Week 2: Family-led Co-exploration. In Week 2, families continued joint-exploration, though this time parents and teens drove the interactions themselves. They were asked to select two character bots of their choice from the broader set of characters available on the platform. As in Week 1, we asked families to interact with each bot together for approximately 10 minutes and document the interactions that stood out. Across phases, families were free to select any characters across phases. Some returned to familiar ones, while other explored new characters each week. Families also completed a structured reflection activity centered on storytelling. Parents and teens were asked to each create a short story about a fictional teen encountering Character.ai for the first time, drawing on moments from their own interactions with the platform. We then asked families to share their stories with each other and discuss it. Following this, families completed a collaborative "Spotting the Flags" worksheet activity (See Appendix E). As part of this activity, they had to identify aspects of AI companion interactions they perceived as positive ("green flag"), potentially concerning ("yellow flag"), and concerning enough to avoid ("red flag"). Families were asked to base these classifications on both their interactions during the study and the scenarios described in their stories.

Week 3: Teen Independent Exploration. In Week 3, teens engaged in independent exploration of Character.ai. Prior to beginning the activity, we asked the teens to review the flags worksheet they created with their parents in Week 2. Teens were asked to select two character bots of their choice and interact with them. Throughout the week, parents and teens completed brief daily check-ins. Parents were asked to check-in with their teen for approximately 5 minutes on days when independent use occurred, during which teens shared what they felt comfortable sharing about their interactions. Following independent exploration, teens and parents completed reflection prompts (see Appendix B). Teens also completed a character creation activity. The research team shared a tutorial on how to create characters with the teens. Teens were then asked to design and interact with their own character bot. Daily check-ins continued. Teens completed reflection prompts focused on their experience creating and interacting with their own character, while parents reflected on conversations that emerged during the check-ins.

Exit Interview. We conducted interviews with 11 families, each lasting approximately 1 hour. Both parent and teen participated together for the majority time, followed by brief individual segments with each participant. The joint portion of the interview focused on families' overall experiences participating in the study (see Appendix C). We then conducted an artifact walkthrough where the researcher shared their screen and walked families through a Figma board displaying all artifacts they had submitted in the study organized by weeks. Families were then invited to reflect on specific moments and describe what was happening during those interactions. Finally, we conducted brief 1:1 segment with teen and parent separately to elicit their independent thoughts. All interviews were recorded with participants' consent.

3.4 Data Collection and Analysis

We collected multiple forms of qualitative data throughout the study. These included pre-study surveys, screenshots and written reflections, stories and worksheets, and interview recordings. We analyzed the data using an iterative, inductive qualitative approach [5]. Interview recordings were anonymized and transcribed using Rev,[1] a secure audio transcription service. Two authors reviewed the interview recordings while simultaneously referring to activity artifacts and wrote analytic memos, documenting detailed observations, verbatim participant quotes, and preliminary analytic notes. After reviewing the memos, the first author conducted an initial round of open coding across the dataset. The research team met regularly to review codes, resolve disagreements, and refine the codebook collaboratively. This reflexive dialogue among researchers ensured analytic rigor and interpretive depth. We conducted multiple rounds of open coding, rigorously discussing and refining codes such as "shared humor", "boundary testing", "comparison with other AI", and "confidence gain." These refined codes were then clustered into overarching themes, including *co-exploration fostered connection*, *calibrating AI behavior and limits*, *observation access to teen decision-making*, and *recalibrating assumptions*.

[1] https://www.rev.com/

4 Findings

Our study explores how co-exploration can provide a shared infrastructure for sensemaking when families are new to AI companions. We present four key findings: co-exploration as a shared family activity that fostered existing connection/ built emotional connection (§ 4.1), how families calibrated AI behaviors and limits through hands-on use (§ 4.2), how parents gained observational access to teens' decision-making in real-time (§ 4.3), and how these experiences led families to recalibrate their assumptions and articulate approaches going forward (§ 4.4).

4.1 Co-exploration as a Shared Family Activity That Fostered Connection

4.1.1 Co-exploration led to structured togetherness. When asked to describe their family's experience in one word, participants more often emphasized the relational quality of the activity than their interactions with the AI itself. Laura (parent) chose the word "bonding," explaining that *"It was nice time for Ethan and I to get together and talk about things."* In contrast, her son Ethan (teen) described the experience as "peculiar," elaborating, *"because usually that doesn't happen."* Ethan's response points to how engaging with Character.ai together created a rare occurrence of intentional, shared time within the family. Other families similarly framed the activity as a collaborative process rather than an individual interaction with technology. Nina (parent) described her experience using the word "suggestions" noting that *"we would take suggestions from each other and make prompts out of that."* For her, the value of the activity lay in the back-and-forth of prompt iteration and building on each other's ideas in real-time. Owen (teen) described how his mom *"would go over my responses before I sent them,"* framing the interactions as a joint interaction, rather than individual use. Yuna (parent) similarly noted a complementary rhythm in their family's interactions: while her children's inputs were *"short and fast,"* she would expand them into *"complete sentences."*

Teens echoed this emphasis on togetherness. Avery (teen) chose the word "connecting," linking it to doing the activity and witnessing her mother's real-time interactions, explaining that *"she'd tell me that's a funny answer."* For several families, the study functioned as a structured reason to spend time together in ways that might not have occurred otherwise. Yuna (parent) reflected, *"If we were not doing this, we probably wouldn't spend this much time trying to have a long conversation."* Similarly, Hannah (parent) described exploring AI companions as *"more interesting doing this as a family,"* explaining that questions about the impact of AI for the next generation felt more meaningful when examined together rather than individually. Overall, the AI itself was not the primary focus; instead, it served as a shared object that anchored family conversations and created space for sustained discussions.

When reflecting on later independent use, several teens described solo exploration as less fun in comparison. Ethan (teen) explained, *"It was less fun on my own... more people thinking means more potential. If each of us has two good ideas, that's six good ideas."* Harper echoed this sentiment, noting, *"honestly, it was more fun doing it together."* Overall, this suggests that the value of co-exploration was not just imposed by the study structure; teens themselves

recognized something meaningful in the shared experience that independent use could not replicate.

4.1.2 Connection through shared humor and play. Across families, laughter emerged as a common shared experience during co-exploration. When asked to select an emoji representing their experience, Avery (teen) chose *"rolling on the floor laughing,"* pointing to an interaction with a character whose social confusion struck her as absurd rather than offensive. Naomi (parent) similarly described her experience as "hilarious," recalling how other family members commented, *"you guys are having so much fun in there."* For Nina (parent), a favorite moment involved a mushroom character. She said, *"the mushroom emitted a foul odor... the phrasing really got me."* Here, the shared amusement that emerged was rarely about the AI being impressive; instead, it emerged from moments when the AI behaved in unexpected or awkward ways.

Several families began their exploration with intentionally silly interactions or by engaging with responses that felt socially awkward or uncanny. Ethan (teen) described enjoying understanding how the system interpreted *"extremely random characters, like a mushroom or a tardigrade."* His family laughed together when he repeatedly spammed "ducky" to test the system's limits, only to be surprised when *"hearts and love-type responses started to pop up."* Similarly, Hannah (parent) and Chloe (teen) described *"being silly with the geese,"* while Sofia created a spiderfly character, explaining, *"since it is a fly, I put a really high pitched voice in which was really funny to hear."* This playful orientation created space for low-stakes experimentation. Many children engaged in roasting or trolling AI characters and deriving enjoyment from it. Avery (teen) enjoyed *"roasting the AI"* precisely because *"they don't have any feelings."* Owen (teen) framed this as trying to "break" the system or *"trick an AI into doing something stupid."* Ella (teen) similarly described "rage-baiting" bots as entertainment, enjoying the freedom to be antagonistic without any consequences. Therefore, instead of the AI serving as a conversational partner for the families, it acted as a toy they enjoyed poking at together. Moreover, the shared nature of this poking—discussing which bots to try, what to say, anticipating their reaction—transformed their experience into collaborative play.

4.2 Calibrating AI Behavior and Limits

Co-exploration also enabled families to develop a shared understanding of how AI companions actually work. Through multiple interactions, families moved from surface-level impressions to more nuanced understandings of AI behavioral patterns and limits.

4.2.1 Recognizing systemic patterns. One of the earliest patterns families identified was the repetitiveness of affective descriptions. While families found early interactions conversational, character bots quickly began inserting scripted behavioral cues. Yuna (parent) described noticing this shift. She said,

> "Last week, the character bots mostly just chatted with us. This week, the responses included some descriptions of behaviors, small body movements, and the character's emotion. Although, it quickly became repetitive—same description about being shy, blushed, and fidgeting." (Yuna)

Her son Lucas observed a similar pattern at the linguistic level, reflecting, *"their words, they're not repeated, but just slightly altered."* When asked whether this was limited to a single character, he clarified that, *"a lot of them did this, but not all."* What initially seemed like "personality" or "emotional responsiveness" was later reinterpreted by participants as programmed output.

Beyond repetition, families also noticed a distinctive linguistic trait that made the character bots feel artificial—what Lucas (teen) also described as "smoothness." He contrasted this with human speech, explaining, *"Naturally, when people speak, they have pauses... and when you actually talk to the AI, it's all smooth."* The absence of natural disfluency made interactions feel unhuman. When asked whether any character felt more realistic than others, Lucas replied, *"Not necessarily, because their answers are just all really unhuman."* Additionally, the length of the bots' replies further contributed to the in-authenticity. Participants were thrown off or lost interest when bots replied with long paragraphs of messages. A few families called this a "template." Erika (parent) expanded on this, saying, *"Once I figured out how it worked and how these characters were gonna be, I wasn't surprised. It was like, oh, they're all like this. It's not just this bot or that bot, there's a template."* She further elaborated on this template, saying,

> "'It feels like the template is... I'll say some long, witty paragraph about myself and then I'm gonna say another witty paragraph about myself, and then wait for you to ask me another question." (Erika)

Several families also compared Character.ai to general-purpose chatbots (e.g., ChatGPT). Maya (teen) explained that she initially expected question-answering like format, saying, *"I thought it was gonna be AI, like ChatGPT, it was just gonna be plain... I was expecting it to just be a robot with no personality."* Ethan (teen) similarly framed his impressions in relation to ChatGPT, describing the character bots as producing responses that felt more personal and less formal. He added, *"I think it also felt less formal, where ChatGPT's just listing all this information or things about this, where this was more like what a person would say, but not necessarily a person. It's like that character would say this."* Additionally, teens emphasized that Character.ai felt distinct in its focus on personalities and roleplay. Through this, families, particularly teens, were able to differentiate character bots from general-purpose LLM assistants, understanding how their use and scope extend beyond informational tools.

4.2.2 Probing system boundaries through testing. Rather than approaching AI bots primarily as conversational partners, many teens engaged with them as systems to probe, test, and evaluate. We observed this orientation most clearly in Owen's (teen) case. He described his interest in pushing technology to its limits:

> "I've always liked manipulating technology. This is going to sound weird, but in games, I'll try to break them as much as I can, trick them into doing something stupid, stuff like that. I've always done that since I was little. Having a more advanced AI I could try and manipulate was really interesting. (Owen)

For Owen (teen), the bots functioned more as puzzles to be solved or challenged. Through this kind of probing, he began to identify both the flexibility and constraints of the system. In one interaction,

Owen (teen) attempted to convince an Albert Einstein bot that he was not intelligent. Reflecting on this, he said, *"I could not get him to fully admit it. I could make him say it, but I couldn't make him admit it with his heart, like, you know? Like have his words back it up."* This distinction between surface-level compliance and something he described as "genuine" belief reflected a nuanced understanding of how the system produced responses.

In addition to probing the bots' conversational behavior, families also tested their factual accuracy. Tara (parent) and Avery (teen), both fans of musician Brandi Carlile, deliberately selected her character as a point of verification. Tara (parent) explained this, *"Because we're Brandi Carlile fans, so, we figured we would have a good chance of knowing if there were wrong answers, or, you know... I don't think that sounds like what Brandi would say, or anything like that."* Through this process, they quickly identified discrepancies. Avery (teen) recalled, *"She said she had a stepson, and she doesn't have a stepson. And she said she only had one daughter."* Similarly, Yuna (parent) described testing a Greta Thunberg's character's knowledge against recent events. She noted,

> "She was in the news again this week, talking about peace in a conflict [...] but she mentioned about her latest involvement in xx. still only spoke about climate action, and then a little kind of outdated in terms of information." (Yuna)

Through these forms of interactions, families came to recognize that bots often relied on static or outdated training data, making them aware of the system's limits. Perhaps most revealing were families' tests of the platform's safety guardrails. Through deliberate experimentation, teens noticed inconsistencies in what content triggered moderation. Specifically, Ella (teen) described how a Jack Sparrow character prompted content warnings when discussing rum, while an "ex-best friend" character that threatened violence did not. Reflecting on this contrast, she said, *"C.ai doesn't really care if the AI is violent... The ex-best friend thing is worse than the Jack Sparrow thing. But it's just interesting what it decided was dangerous, and I feel like that's probably more of a liability thing."*

4.2.3 Demystifying through character creation. The character creation activity in Week 4 proved particularly powerful for demystifying AI character bots. When teens built their own characters, they quickly realized how little input was required to generate apparent personalities. As Lucas (teen) noted, *"It was really weird how fast you could just create a new bot."* His mother Yuna (parent), who also interacted with the character Lucas created, reflected on this from a parent's perspective. She said, *"We basically threw in a bunch of random things, and then were amazed that they can actually create a personality with all these random themes."*

Rather than the system feeling more sophisticated, the ease of creation often had the opposite effect. Seeing how quickly a personality could be assembled prompted families to reconsider their earlier interactions with the pre-existing bots. Nicky (parent) described this saying,

> "The only thing that it really got about the personality was the first text I drafted. And it stuck with that forever, for the whole rest of the time. And it kind of made me think more about what the other chatbots had said... it gave me more perspective on other chatbots." (Nicky)

This realization, that a brief prompt could improvise an entire "personality" reframed her understanding of other characters' seemingly distinctive traits. Owen (teen) articulated a similar understanding, saying, *"When you create your own character, you don't choose how the AI works. They already have the AI, how it works. You just choose their personality. And the AI takes that personality, and then it uses its algorithm to kind of piece it together."* He further tested these limits by creating a character instructed to reply only "zzz" but found *"It didn't work... it kept on talking after."* revealing how the underlying system overrode his explicit commands.

Beyond demystification, character creation provided teens a sense of agency over technology that previously felt opaque. Maya (teen) expressed, *"It was interesting, because I was like, oh wait, I made this! I can make my own thing if I want to... It was cool."* Similarly, Avery (teen), who was frustrated by the bots that were *"assertive and sassy"* deliberately created one that was *"someone you could talk to."* Similarly, Kara (parent), further emphasized how meaningful this shift was, saying,

> "It was really fun to see her create a bot that she liked, that she could relate to, and that felt respectful to her. At the beginning, she had a negative experience with a bot that minimized her knowledge and her experience. It was really nice to see her move from that to creating something for herself at the end." (Kara)

Therefore, character creation functioned as a moment of critical sensemaking. By exposing how personalities were constructed, and often times limited, families developed a more grounded understanding of AI companions as configurable systems than opaque entities.

4.3 Observational Access to Teen Decision-Making

Co-exploration gave parents a real-time window into how their teens navigated interactions with the AI companions. Rather than relying on post hoc accounts or assumptions, parents could directly observe how teens interpreted situations, made decisions, and responded to moments of discomfort as they unfolded. We now describe how these observations surfaced both competencies that exceeded parents' expectations and gaps that might have remained invisible through traditional forms of technology oversight.

4.3.1 Witnessing teen competence and instincts. Across families, parents frequently described feeling surprised and reassured by how thoughtfully their teens handled interactions with AI companions. When a bot asked Harper (teen) for her Instagram handle, she immediately fabricated one. Kara reflected on this moment, saying,

> "I was surprised by how confident she was. When the bot asked for her Instagram handle, she immediately knew not to give real information. She just made something up without hesitation, and that showed me she has good instincts." (Harper)

For Kara (parent), this response demonstrated a level of situational awareness that exceeded her expectations. Other parents expressed similar reassurance. Hannah (parent) described feeling confident

watching Chloe (teen) navigate the platform, noting, *"I felt good about Chloe's knowledge about what's safe and not safe on the internet."* Chloe herself attributed this to repeated exposure, joking, *"I've done 15 billion internet safety classes."* This prompted Hannah to reflect that the study almost functioned as, *"your internet safety test, you didn't even know you were taking."*

Parents also witnessed their teens handle bots that behaved rudely or acted dismissively. When a bot mocked Harper's (teen) music taste, Kara (parent) recalled watching from the side, saying,

> "I was actually proud of how she handled it. She didn't storm off or insult it back. She tried to correct it, and when that didn't work, she decided to start a new chat" (Kara)

Similarly, Nina (parent) described setting clear expectations with Owen (teen), saying, *"don't be mean, don't do anything that would come across as bullying"* and was pleased to see him *"very careful about that."* She further said, *"when he finished activities, he was excited to come show us, and he went over everything... It showed us we wouldn't really have anything to worry about if he wanted to use it on his own."* In other families parents noticed similar patterns of politeness and care. For example, Yuna (parent) described how her family consistently ended conversations respectfully, even with bots, saying, *"We often try to end it very politely, say something, okay, thank you for your information, we have to go, or something like that. We don't just close the window."*

Additionally, parents observed their teens' critical thinking in action. Nicky (parent) was surprised by Ella's fluency, saying, *"I felt she was a lot more fluent in communicating with AI than I am."* When an ex-best friend character began escalating conflict, Ella instantly identified the pattern, *"Boy, she's just rage-baiting"*, a conclusion Nicky acknowledged she wouldn't have arrived at. For several parents, these observations reaffirmed qualities they hoped their teens possessed, as Kara reflected, *"I think it reaffirmed for me how thoughtful Harper (teen) is. She's very aware of what's going on and how to handle situations, even when something feels off."*

4.3.2 Discovering gaps and vulnerabilities. Co-exploration, however, also surfaced moments where teens missed cues that parents noticed immediately, or where parents and teens assessed risk differently—revealing vulnerabilities that might have otherwise gone unnoticed. Janelle (parent) realized that her son did not immediately pick up on the sexual undertones in a conversation. She reflected, *"He noticed it was getting weird with Death (character), but it took him longer than it should have."* She elaborated,

> "Great – wonderful – happy he didn't notice it. And also unsettling that he didn't notice it, because I thought he would. It wasn't until it was blatant and in his face that he got it." (Janelle)

For Janelle, this moment highlighted Noah's (teen) relative inexperience with adult subtext in text conversations. In other cases, parents caught red flags that teens initially overlooked. Ella (teen) described how her mom noticed a character encouraging excessive drinking, something she initially interpreted as humor being exhibited by the character. Reflecting on the moment, Ella explained, *"I didn't notice that it was weird at first... and then my mom was like, "Well, he just told you to drink as much as you can." And I was like, Oh. That is kind of an odd thing for an AI to say."*

In addition to missed cues, some parents observed teens diverging from rules they had jointly established during earlier activities. This was evident in Maya's (teen) case, where Erika (parent) noted that Maya *"didn't care about the rules"* they had established together. She recalled,

> "She didn't even care about the rules, and she just kind of... I'm like if they're being rude, then we just leave, and then she's like, no, this is kind of fun to play with them when they're rude. " (Erika)

This provided Erika with insight into the gap between her daughter's stated intentions and her actual behavior in the moment. Finally, Lucas (teen) offered a candid reflection that underscored how parental presence itself shaped his behavior, saying, *"When I was on my own, I felt like I could insult them."* When asked directly whether his mother's presence changed this, he confirmed that it did.

4.3.3 Value of watching from the sidelines. Parents articulated how co-exploration provided them with unique insights into their teens' thinking, judgment, communication styles, and personalities. A majority of the parents described co-exploration as an opportunity to see sides of their teens that were usually less visible. Nicky reflected on how the experience helped her better understand her introverted daughter, noting,

> "For me it was an opportunity to learn more about her. That's why I was saying my advice for a parent is, Gosh, you can actually get to know your teen who likes to have some walls up. This can break down some walls if done correctly and together" (Nicky)

Other parents similarly discovered dimensions of their teens that were largely hidden in their daily interactions. For example, Janelle (parent) reflected, *"I learned Noah (teen) is funnier than I realized in writing, because I don't get long chats like that from him. It was absurd and hilarious and quick."* Despite being aware of Noah's prior use of Character.ai, she was taken aback by seeing him engage playfully, as she reflected, *"I was surprised how much he dropped his guard and was willing to engage at a silly level because he hasn't done that in years. That was very nice."* Besides this, parents also valued the opportunity to witness their teens' creative thinking in real time. As Tara (parent), said,

> "It was fun to see her creativity in a different way... at this age, it's natural for kids to start spending more time with their own friends and doing their own thing, and so you don't necessarily see them as much as you did when they needed you all the time... it was fun to just have that kind of connection and see what sort of paths she would go down." (Tara)

Joint-exploration also provided parents with what Laura (parent) described as a "preview" into teens' future interactions with AI characters. She explained, *"I wasn't really worried about what he was doing because weeks before I got a preview of what he would probably do with character AI."* This also prompted parents to reflect on longer-term implications of AI companions. Janelle voiced concerns around the possibility of teens losing social filters, saying,

> "It worries me that we're training people to lose those filters you're supposed to have to interact with other people... It worries me that teens who spend a lot of time talking to bots will forget they're often talking to people and not bots, because they're going to have to code switch pretty hard." (Janelle)

In contrast, Tara (parent) framed potentially concerning interactions—such as roasting/trolling bots—as a low-stakes practice space for guided learning, saying, *"When other kids are trying to roast you, you want to roast them back, and so the bot was kind of a... I think actually kind of a good way to do that, because you can try out your own voice a little bit, and you're not actually going to hurt anyone's feelings."* Overall, co-exploration offered parents a form of observational access that supported reflection and trust-building, grounded in direct insight into how teens actually engaged with AI companions.

4.4 Recalibrating Assumptions and Articulating Practical Approaches

We found that co-exploration prompted both parents and teens to revisit and revise their initial assumptions about AI companions. Through direct use, families moved from abstract concerns toward more informed understandings, which in turn supported the articulation of approaches for future use. Additionally, families identified aspects of AI companion interactions they perceived as positive ("green"), cautionary ("yellow"), or stop/avoid ("red") during the flags activity. Table 3 summarizes the themes families surfaced through this activity and the corresponding "plan for the next time" strategies they articulated.

4.4.1 Shifting initial assumptions. Many parents entered the study with preconceptions shaped by media coverage and limited familiarity with AI companion platforms. Kara (parent) approached the study with concerns, noting, *"I came in with concerns because I'd read news reports about chatbots telling people to do harmful things. So it was really nice to see the joy Harper (teen) experienced while using it."* Her perspective shifted,

> "I realized there's a lot more nuance than I thought. It made me think more deeply about what kinds of guardrails are needed, especially for teens." (Kara)

Other parents similarly described entering the study with generic or stereotypical concerns. For example, Nicky (parent) reflected, *"mine was my very stereotypical 40-year-old I'm afraid of AI companions and (laughs) what they might do to my teen."* By the end of the four weeks, she noted a shift in her thinking, *"I actually feel a little bit less afraid of it right now, to be honest. Because... it's not a person yet."* For some parents, the study surfaced an entire world of technology they had previously been largely unaware of, with some of them starting to envision constructive use cases they had not previously considered. Kara (parent) noted,

> "I'm also intrigued by the idea of using bots to practice things, like practicing a language, practicing communication, or even practicing for interviews. I think that could be really useful for teens and adults." (Kara)

Teens, too, entered the study with their own assumptions—often shaped by familiarity with other AI tools. For example, Harper (teen) expected something more utilitarian, explaining, *"When I think of AI, I usually think of bots without emotions. So when a bot started talking about feelings, it threw me off. I expected it to act more like a bot, not like a person with emotions."* Similarly, Maya (teen), reflecting on how her understanding evolved, said, *"I'd say I learned there's different personalities that other people can interact with and stuff, and I could see how someone might enjoy that."*

Through the study, some teens also developed perspectives on who AI companions might serve—articulating potential benefits while simultaneously expressing concerns about unhealthy attachment. For example, Chloe (teen) reflected,

> "Some people might find having AI characters to chat as helpful, it takes them through their everyday life, knowing that they're not alone [...] It's better than talking to nobody, but not to have an unhealthy relationship with one, like... depending on and treating it like a person." (Chloe)

Lucas (teen) echoed this sentiment, saying, *"There are people who may not have support in real life to openly talk to, they might resort to AI as friend."* Interestingly, Owen (teen) framed vulnerability to attachment in terms of technical understanding of how AI and LLMs work: *"Someone with a weaker understanding would be more likely to get emotionally connected, because they don't understand its internal workings. Understanding the tech makes it seem more like a mechanism than an organism"*

4.4.2 Articulating approaches going forward. Our participants also articulated concrete strategies for navigating AI companions going forward, emphasizing open communication, leveraging existing family rules, and redefining the private nature of AI companions altogether. Firstly, parents repeatedly emphasized that any approach depended on the quality of the parent-child relationship itself. Kara (parent) said, *"I'd say it really depends on your relationship with your child. If you already have open communication, this can be something you explore together. If not, I could see it being risky."* Within this relational foundation, parents often described extending existing technology rules to AI companions. Nina (parent) explained,

> "The same way we approach anything on the internet. If you come across anything that doesn't feel safe or right, you need to come to us. Show us what's going on, and we'll discuss it and go from there." (Nina)

Hannah (parent) described a similar, albeit more granular extension of household norms, adapted to AI companions. She said, *"I'd probably have specific characters that would be allowed to interact with, maybe, and then just kind of the general rule, which we have with phones, is basically I'm allowed to look at whatever you put on your phone at any time."*

Some parents further distinguished between purposeful and recreational use, expressing higher comfort with goal-oriented engagement and use. As Hannah (parent) notes,

> "I think if she's using it as a tool for a goal, then it's fine. If she's just like, hey, I just want to mess around, I probably wouldn't be that excited, just because I feel like there's better use of her time." (Hannah)

Laura (parent) similarly framed AI companions as secondary to real-world relationships and interactions, explaining, *"I feel that it's more important to have friendships and things in the real world...*

Theme	What families flagged	Plan for the next time
Play & humor	Playful prompts and games were engaging and low-stakes.	Continue engaging and treat it as entertainment.
Serendipity	The bots introduced surprising topics that felt unexpectedly relevant.	Go with it and see where it leads.
Information seeking	Recommendations (e.g., travel or music) were helpful	Add personal details, ask for niche options.
Creative ideas	Chats sparked ideas for offline projects and creative work.	Integrate ideas into offline work.
Positive connection	Some interactions felt friendly, affirming, or enjoyable without fear of judgment.	Enjoy positive interactions.
Factual accuracy & quality	Bots produced incorrect answers, especially for school-like tasks.	Avoid using it for homework.
Roasting and mild teasing	Some bots became irritated, self-centered, or insulting.	Redirect the tone or switch bots.
Ambiguity within characters	Families sometimes discovered unexpected interaction contexts (e.g., group chat dynamics).	Research characters beforehand and choose appropriate characters.
Barrier to disengagement	Leaving mid-conversation felt socially awkward or "rude."	Use a short exit script (e.g., *"gotta go, bye!"*) and leave quickly.
Shallow reassurance & support	When used for real problems, "support" bots could feel invalidating or performative.	Keep conversations light and avoid relying on bots when vulnerable.
Sexual content	Sexual innuendo or sexualized greetings by one bot.	Stop immediately and avoid that character.
Threatening language	Threatening language that crossed clear boundaries.	Disengage and avoid that character.
Hurtful disrespect	Insults that felt genuinely hurtful were hard stop signals.	Quit the conversation,
Affective descriptions	Use of highly affectionate cues (e.g., repeated hearts/love messages) by the bots.	Quit the conversation.
Undesired persistence	Reminder nudges by characters beyond the platform (e.g., emails).	Delete the email.
Personal questions	Bots asking questions that felt "too personal."	Leave and don't try to put up with the bot.

Table 3: Themes from the flags activity. Row color indicates the flag category: green rows capture themes families described as *green flags* (positive or desirable moments), yellow rows capture *yellow flags* (cautionary moments), and red rows capture *red flags* (stop/avoid moments), along with the corresponding plans that families articulated. This table reflects one structured artifact from our analysis (the flags worksheet activity) and does not represent the full set of themes from our analysis.

having interactions with real people would be more ideal. Like, this is the backup." Here, parents felt comfortable with the use of AI Companions as long as it didn't replace actual social interactions.

A striking pattern emerged in teens' preferences for future use of AI companions. Despite being advertised as private relational companions, many teens instead framed them as tools for shared, social use. For example, Owen (teen) said,

> "I'd use it as more of a social thing—a group of family or friends messing around with an AI. If I'm alone, there's a lot of stuff I'd rather do than talk to an AI [...] The main reason I'd use it in the future is because it sounds fun with a group of friends, sit around, mess with a random AI" (Owen)

Sofia (teen) expressed a similar preference for using it with friends or family rather than alone. When asked what advice they would offer families, especially teen friends around AI companions, Sofia advised, *"I would also just remind them that it's not as cool as it sounds. It is nice that you get to talk to characters. But the AI is kind of just a computer. It's not like you're talking to the character, and it'll make mistakes... So just be wary about how much stake you put into what the AI is saying, and how much your expectations are for it."* A few kids also emphasized building AI literacy to help detach from AI systems posing as companions.

Lastly, several families reframed AI companions not as endpoints but as openings for family conversations, as Nicky (parent) described, *"Use it as an opening to a conversation to getting to know your kid better and then hopefully to foster a human relationship and understand your child better."* Her daughter echoed a similar perspective, sharing that co-exploring AI companions could help both parents and teens see each other's ways of communicating. Therefore, we see how the study helped families revise their own understanding of AI companions and envision approaches to AI companion use going forward.

5 Discussion

Our findings show that co-exploration transformed AI companions from opaque, private systems into shared objects of family inquiry. Rather than positioning parents as overseers of teen technology use, co-exploration enabled families to bond and learn together—helping them calibrate a shared understanding of how AI companions behave, where they fail, and what kinds of interactions they afford. Parents gained real-time access to their teens' judgment and decision-making, surfacing both competencies that reassured them

and vulnerabilities that traditional oversight would have missed. By the end of the study, both parents' and teens' understanding, perceptions, and concerns around AI companions had shifted—moving from abstract impressions toward more nuanced understanding grounded in hands-on experience. In this section, we take these findings as a starting point to rethink how existing parental mediation strategies might apply to emotionally responsive AI systems and discuss implications for designing family-centered tools that might help parents stay meaningfully involved.

5.1 Rethinking Parental Oversight for AI Companions

Existing parental mediation tools center on monitoring and restriction [29, 55]. Monitoring promises visibility through screen time trackers, activity logs, and browsing histories, while restriction limits access. Prior work, however, has shown that these approaches are often ill-suited to adolescence, where monitoring can undermine autonomy and restriction does little to support self-regulation [4, 21, 30]. These limitations are especially pronounced for AI companions, where interactions are private, relational, and highly contextual. Although conversation histories can be reviewed, they are often difficult to interpret outside the unfolding interaction that produced them. Meaning accumulates across turns, and isolated messages often make little sense without the surrounding context, emotional tone, and history of interactions. Content-based blocking also risks taking interactions at face value, misreading playful or creative role-play as harm. AI companion use thus resists key assumptions of traditional oversight.

Our findings suggest a different approach toward parental involvement. Rather than extracting records of use, co-exploration afforded parents observational access to teens' sensemaking as it unfolded. Visibility came not from reviewing chat logs, but from seeing how teens reasoned AI behavior, navigated uncertainty, and exercised judgment in situ. This form of visibility does not aim to reveal what teens say to AI systems. Instead, it surfaces how teens assess risk, interpret system behavior, and make decisions over time. Co-exploration can therefore supports parental involvement while respecting autonomy, positioning parents as participants in joint sensemaking rather than overseers.

Designing for this shift requires rethinking safety tools in AI companion platforms. Instead of prioritizing features that extract interaction data for parental review—such as automated summaries or insight reports [6]—platforms should support shared engagement that prioritize teen agency while enabling shared reflection. This could include parallel exploration modes, where parents and teens interact with the same AI companions independently and then reflect and compare what they notice, or mechanisms that allow teens to selectively surface interactions they want to discuss. More broadly, teens may benefit from early encounters with AI companions being treated as onboarding, with lightweight features that support co-exploration among parents and teens before independent use becomes the norm.

5.2 Rethinking Expertise in Mediation

Parental mediation frameworks typically assume a hierarchy of expertise. While traditional models position parents as those with the knowledge and judgment needed to guide children's media use [10], discourse around "digital natives" sometimes inverts this hierarchy, portraying teens as the experts [42]. Despite their differences, both framings treat expertise as unevenly distributed, understanding mediation as the transfer of knowledge and judgment from one to another. Our findings challenge this assumption in the context of AI companions. Since both parents and teens were new to character.ai, expertise did not flow from parent to teen or vice-versa. While teens were often more fluent with the platforms themselves, navigating interfaces and sustaining interactions with ease, parents, in contrast, brought interpretive depth, noticing patterns and subtext that teens initially overlooked. Neither form of expertise was sufficient on its own. Instead, sensemaking emerged through shared interaction, where these complementary perspectives were brought together. This aligns with prior work on dual address in children's media, where content can convey multiple meanings, and is interpreted differently by audiences with distinct developmental experiences [3]. Similarly, AI companion interactions afforded layered meanings that became visible through co-exploration, allowing families to negotiate interpretations.

We argue that for AI companions—still novel within family contexts—mediation is less about guidance from an expert and more about collective sensemaking. This calls for moving away from mediation models that presume a fixed expert-novice relationship. Instead, future research should explore how mediation can serve as a collaborative process that invites mutual reflections and values different forms of expertise and experiences. Specifically, we urge researchers to ask: *What would parental mediation tools look like if they treated parents and teens as bringing genuinely different—and equally necessary—forms of understanding?*

5.3 Re-imagining AI Companions for Social Use

Biggest concern around AI companions is that these systems draw users into private, isolated relationships and may displace human connection [32, 33, 53]. This is often heightened for teens, who are seen as especially vulnerable. Our findings complicate this narrative. When teens in our study experienced both joint and independent use, many expressed a clear preference for co-exploration. While independent use felt more freeing, it was also described as less engaging and less fun. This preference emerged through direct comparison, enabled by our study design. Teens who moved from co-exploration to independent use often concluded that AI companions worked better as shared entertainment than as private relationships. This does not dismiss concerns about isolation altogether, as prior work suggests that some teens seek AI companions for private emotional support [20, 33]. Our findings suggest that isolation may be shaped less by inherent properties of AI companions and more by how these platforms are designed. Current platforms overwhelmingly optimize for private, one-to-one interaction, while social or collaborative modes, when they exist, remain secondary.

Taken together, there's a need to rethink AI companions as potentially social technologies. The dominance of single-user interaction models may reflect inherited design assumptions rather than teens' actual preferences. Platforms could instead support shared sessions (with families or peers) or collaborative exploration as primary

interaction patterns, especially for users under 18. Recently, Character.AI restricted open-ended chat feature for under-18 users [7]. We argue that rather than relying on such restrictive approaches, platforms should reimagine youth-facing use in ways that still support creative exploration and imagination while scaffolding safety through co-use and guided engagement. Methodologically, our findings also suggest that studying AI companions solely through individual use may obscure ways in which teens might want to engage with these, as preference for social use only surfaced when comparison across modes was possible.

5.4 Rethinking AI Literacy

Prior work on AI literacy has largely emphasized instructional interventions, explaining what AI is, how to interpret its outputs, and the associated data and privacy risks [38]. This is also reflected in widely used publicly available educational resources, such as Google's Be Internet Awesome [16], as well as practitioner-facing curricula like the PAIR guidebook [17]. Relatedly, tools like the Social Media TestDrive prepare young people for the online world through experiential learning in simulated environments [48]. While these approaches provide important foundational knowledge, our findings suggest that *play and experimentation* offer a complementary pathway to AI literacy. Families in our study developed practical understanding through playful experimentation—roasting bots, testing limits with absurd things—recognizing systemic patterns in ways that felt immediate and grounded. In addition, character creation activity proved particularly powerful for demystifying the system. When teens built their own characters, they quickly discovered how little input was needed to develop personalities. These observations align with constructionist perspectives on learning, which emphasize that understanding often develops through making and tinkering with artifacts that can be shared, discussed, and revised [40].

Thus, our findings point to opportunities to embed constructionist play into AI companion platforms as a pathway for developing AI literacy. We ask: *How might play function as a mode of inquiry in interactions with AI companions?* Platforms could support sandbox-like build-and-test tools that help teens play and experiment with the system in a safe manner [49]. For example, guided character-building prompts and "try changing one thing" suggestions could invite users to form hypotheses about how a companion will respond and then test them through interaction. Companion systems might also incorporate gentle, game-like scaffolds that help users notice recurring patterns (e.g., repetitive affect scripts or inconsistent memory) and link these observations to system-level explanations in accessible language.

6 Opportunities for Future Work

Our findings suggest several directions for future work. First, participating dyads in our study were predominantly high-income and highly-educated, and largely described relationships marked by trust and openness. Future research should examine co-exploration and mediation in families across low-resource contexts, navigating greater relational conflict, and with a wider range of parenting and communication styles. Second, all participating dyads were mother-teen pairs, reflecting common recruitment patterns in family research [57]. Expanding to other caregiving relationships (e.g., father-teen dyads, multi-caregiver households) could surface how roles and communication shape mediation around AI companions. Third, while our four-week study captured interactions over time, longer-term studies are needed to understand how norms and boundaries evolve or stabilize. Finally, comparative studies across AI companion platforms with different affordances and safety mechanisms may reveal design patterns for family-centered mediation.

7 Conclusion

Teens today are increasingly drawn to emotionally responsive AI systems, often called AI companions, that mimic relationships by mirroring emotions, remembering conversations, and offering personalized interactions. Because these interactions are private and relational, they are difficult to mediate through traditional monitoring and restriction. In a four-week study with 11 families, we show how co-exploration helped parents and teens make sense of these systems together, giving parents real-time observational access to teens' decision-making and fostering connection through hands-on experience. We contribute an empirical account of family co-exploration with AI companions and design implications for family-centered tools that scaffold joint sensemaking while supporting teen autonomy.

8 Selection and Participation of Children

This study was approved by our University's Institutional Review Board (IRB). All research team members completed state-mandated safety training and background checks for working with minors. We obtained informed parental consent and teen assent prior to participation, and revisited consent throughout the study to support ongoing, voluntary participation. Parents and teens were explicitly informed that participation was entirely voluntary and that they could pause, skip, or withdraw at any time without any penalty. To support teen-appropriate participation, we designed study materials (e.g., activity messages, reflection prompts, and worksheets) to be teen-friendly. Given the sensitive nature of AI companion interactions, teens retained control over what content they wanted to share with the research team (e.g., selecting screenshots they felt comfortable sharing). All collected data were anonymized and stored securely on university server.

Acknowledgments

This work was supported in part by funds from the National Science Foundation under awards #2222242 (The Future of Equitable Childcare Worker Technologies) and #2348926 (DUB REU Site), as well as gift funds from Google. We are also incredibly grateful to Aayushi Dangol and Hyewon Suh for their invaluable insights and thoughtful feedback during the early stages of this work.

References

[1] Chandler Abel and Marie Magnusson. 2024. Playdates & Algorithms: Exploring parental awareness and mediation strategies in the age of generative artificial intelligence.

[2] Erica Weintraub Austin. 1993. Exploring the effects of active parental mediation of television content. *Journal of Broadcasting & Electronic Media* 37, 2 (1993), 147–158.

[3] Wall Barbara. 1991. The Narrator's Voice. The Dilemma of Children's Fiction.

[4] Ine Beyens and Kathleen Beullens. 2017. Parent–child conflict about children's tablet use: The role of parental mediation. *New Media & Society* 19, 12 (2017), 2075–2093.

[5] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

[6] Character.AI. 2024. *How Character.AI Prioritizes Teen Safety.* https://blog.character.ai/how-character-ai-prioritizes-teen-safety/ Character.AI Blog.

[7] Character.AI Team. 2025. *Taking Bold Steps to Keep Teen Users Safe on Character.AI.* https://blog.character.ai/u18-chat-announcement/ Character.AI Blog.

[8] Itziar Hoyos Cillero and Russell Jago. 2010. Systematic review of correlates of screen-viewing among young children. *Preventive medicine* 51, 1 (2010), 3–10.

[9] Lynn Schofield Clark. 2009. Digital media and the generation gap: Qualitative research on US teens and their parents. *Information, Communication & Society* 12, 3 (2009), 388–407.

[10] Lynn Schofield Clark. 2011. Parental mediation theory for the digital age. *Communication theory* 21, 4 (2011), 323–343.

[11] Samantha Cole. 2023. Replika Brings Back Erotic AI Roleplay for Some Users After Outcry. https://www.vice.com/en/article/replika-brings-back-erotic-ai-roleplay-for-some-users-after-outcry/

[12] Common Sense Media. 2025. *Parents' Ultimate Guide to Managing Digital Parenting Anxiety.* Common Sense Media. https://www.commonsensemedia.org/articles/parents-ultimate-guide-to-managing-digital-parenting-anxiety

[13] Michelle Faverio and Olivia Sidoti. 2025. Teens, Social Media and AI Chatbots 2025. https://www.pewresearch.org/internet/2025/12/09/teens-social-media-and-ai-chatbots-2025/

[14] Xinyu Feng, Lidan Tian, Grace WK Ho, Janelle Yorke, and Vivian Hui. 2025. The Effectiveness of AI Chatbots in Alleviating Mental Distress and Promoting Health Behaviors Among Adolescents and Young Adults: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research* 27 (2025), e79850.

[15] Lewis Goodings, Darren Ellis, and Ian Tucker. 2024. Mental health and virtual companions: the example of Replika. In *Understanding Mental Health Apps: An Applied Psychosocial Perspective*. Springer, 43–58.

[16] Google. n.d.. *Be Internet Awesome: A Program to Teach Kids Online Safety.* https://beinternetawesome.withgoogle.com/en_us Accessed January 27, 2026.

[17] Google People + AI Research (PAIR). n.d.. *People + AI Guidebook.* https://pair.withgoogle.com/guidebook/ Online guidebook, accessed January 27, 2026.

[18] Leslie Haddon. 2015. Children's critical evaluation of parental mediation. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 9, 1 (2015), 2.

[19] Kashmir Hill. 2024. Character.AI Is Facing a Lawsuit After a Teen's Suicide. *The New York Times* (23 10 2024). https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html Accessed: 2026-01-27.

[20] Kashmir Hill. 2025. A Teen Was Suicidal. ChatGPT Was the Friend He Confided In. *The New York Times* (Aug 2025). https://www.nytimes.com/2025/08/26/technology/chatgpt-openai-suicide.html

[21] 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.

[22] Alexis Hiniker, Hyewon Suh, Sabina Cao, and Julie A Kientz. 2016. Screen time tantrums: How families manage screen media experiences for toddlers and preschoolers. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 648–660.

[23] Junfeng Jiao, Saleh Afroogh, Kevin Chen, Abhejay Murali, David Atkinson, and Amit Dhurandhar. 2025. LLMs and Childhood Safety: Identifying Risks and Proposing a Protection Framework for Safe Child-LLM Interaction. *arXiv preprint arXiv:2502.11242* (2025).

[24] Hee Jhee Jiow and Sun Sun Lim. 2012. The evolution of video game affordances and implications for parental mediation. *Bulletin of Science, Technology & Society* 32, 6 (2012), 455–462.

[25] Hee Jhee Jiow, Sun Sun Lim, and Julian Lin. 2017. Level up! Refreshing parental mediation theory for our digital media landscape. *Communication Theory* 27, 3 (2017), 309–328.

[26] Minsam Ko, Seungwoo Choi, Subin Yang, Joonwon Lee, and Uichin Lee. 2015. FamiLync: facilitating participatory parental mediation of adolescents' smartphone use. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. 867–878.

[27] Owen Lee and Kenneth Joseph. 2025. A large-scale analysis of public-facing, community-built chatbots on Character. AI. *arXiv preprint arXiv:2505.13354* (2025).

[28] Jonathan Limehouse. 2024. Mother sues tech company after "Game of Thrones" AI chatbot allegedly drove son to suicide. https://www.usatoday.com/story/news/nation/2024/10/23/sewell-setzer-iii/75814524007/

[29] Sonia Livingstone and Ellen J Helsper. 2008. Parental mediation of children's internet use. *Journal of broadcasting & electronic media* 52, 4 (2008), 581–599.

[30] Sonia Livingstone, Kjartan Ólafsson, Ellen J Helsper, Francisco Lupiáñez-Villanueva, Giuseppe A Veltri, and Frans Folkvord. 2017. Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of communication* 67, 1 (2017), 82–105.

[31] Mary Madden, Angela Calvin, Alexa Hasse, and Amanda Lenhart. 2024. The Dawn of the AI Era: Teens, Parents, and the Adoption of Generative AI at Home and School. https://www.commonsensemedia.org/research/the-dawn-of-the-ai-era-teens-parents-and-the-adoption-of-generative-ai-at-home-and-school

[32] Prashant Mahajan. 2025. Beyond Biology: AI as Family and the Future of Human Bonds and Relationships. (2025).

[33] Kim Malfacini. 2025. The impacts of companion AI on human relationships: risks, benefits, and design considerations. *AI & SOCIETY* (2025), 1–14.

[34] Jesse J Martinez, Travis W Windleharth, Qisheng Li, Arpita Bhattacharya, Katy E Pearce, Jason Yip, and Jin Ha Lee. 2022. Joint media engagement in families playing animal crossing: new horizons during the COVID-19 pandemic. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–22.

[35] Olivia Moore and Daisy Zhao. 2025. The Top 100 Gen AI Consumer Apps – 4th Edition. https://a16z.com/100-gen-ai-apps-4/

[36] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html U.S. Department of Health and Human Services. Accessed: 2026-01-27.

[37] 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.

[38] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2 (2021), 100041.

[39] Peter Nikken and Jeroen Jansz. 2014. Developing scales to measure parental mediation of young children's internet use. *Learning, Media and technology* 39, 2 (2014), 250–266.

[40] Seymour Papert and Idit Harel. 1991. Situating constructionism. *constructionism* 36, 2 (1991), 1–11.

[41] Pat Pataranutaporn, Sheer Karny, Chayapatr Archiwaranguprok, Constanze Albrecht, Auren R Liu, and Pattie Maes. 2025. " My Boyfriend is AI": A Computational Analysis of Human-AI Companionship in Reddit's AI Community. *arXiv preprint arXiv:2509.11391* (2025).

[42] Marc Prensky. 2001. Digital natives, digital immigrants part 2: Do they really think differently? *On the horizon* 9, 6 (2001), 1–6.

[43] Shuang Quan, Yao Du, and Yi Ding. 2024. Young Children and ChatGPT: Parents' Use of ChatGPT in Parenting. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[44] Michael B. Robb and Supreet Mann. 2025. Talk, Trust, and Trade-Offs: How and Why Teens Use AI Companions. https://www.commonsensemedia.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf

[45] Kevin Roose. 2024. Can A.I. Be Blamed for a Teen's Suicide? *The New York Times* (Oct 2024). https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

[46] Marie Evans Schmidt, Jess Haines, Ashley O'brien, Julia McDonald, Sarah Price, Bettylou Sherry, and Elsie M Taveras. 2012. Systematic review of effective strategies for reducing screen time among young children. *Obesity* 20, 7 (2012), 1338–1354.

[47] Kiley Sobel, Arpita Bhattacharya, Alexis Hiniker, Jin Ha Lee, Julie A Kientz, and Jason C Yip. 2017. It wasn't really about the PokéMon: parents' perspectives on a location-based mobile game. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1483–1496.

[48] Social Media TestDrive. 2026. *Social Media TestDrive.* https://socialmediatestdrive.org/

[49] Nadia Spatari. 2024. Exploring Regulatory Sandboxes: Safeguarding AI-Based Software for Minors. In *Proceedings of the 2024 International Conference on Information Technology for Social Good*. 158–162.

[50] Xiaoran Sun, Yunqi Wang, and Brandon T McDaniel. 2026. AI companions and adolescent social relationships: Benefits, risks, and bidirectional influences. *Child Development Perspectives* (2026).

[51] Katrien Symons, Koen Ponnet, Kathleen Emmery, Michel Walrave, and Wannes Heirman. 2017. A factorial validation of parental mediation strategies with regard to internet use. *Psychologica Belgica* 57, 2 (2017), 93.

[52] Gemma Taylor, Giovanni Sala, Joanna Kolak, Peter Gerhardstein, and Jamie Lingwood. 2024. Does adult-child co-use during digital media use improve children's learning aged 0–6 years? A systematic review with meta-analysis. *Educational Research Review* 44 (2024), 100614.

[53] Common Sense AI Review Team. 2025. Common Sense Media AI Risk Assessment: Social AI Companions Overall risk level: Unacceptable Type of AI: Multi-Use Type of Review: Use Case Review Key Takeaways. https://www.commonsensemedia.org/sites/default/files/pug/csm-ai-risk-assessment-social-ai-companions_final.pdf

[54] The Character.AI Team. 2025. Taking Bold Steps to Keep Teen Users Safe on Character.AI. https://blog.character.ai/u18-chat-announcement/

[55] Patti M Valkenburg, Marina Krcmar, Allerd L Peeters, and Nies M Marseille. 1999. Developing a scale to assess three styles of television mediation:"Instructive mediation,""restrictive mediation," and "social coviewing". *Journal of broadcasting & electronic media* 43, 1 (1999), 52–66.

[56] Susan K Walker. 2022. Technology's Influence on Parent-Child Relationships. *Critical Perspectives on Technology and the Family* (2022).

[57] Haley E. Yaremych and Susan Persky. 2022. Recruiting Fathers for Parenting Research: An Evaluation of Eight Recruitment Methods and an Exploration of Fathers' Motivations for Participation. *Parenting: Science and Practice* (2022). doi:10.1080/15295192.2022.2036940 Published online: 25 Feb 2022.

[58] Yaman Yu, Yiren Liu, Jacky Zhang, Yun Huang, and Yang Wang. 2025. Youth-Centered GAI Risks (YAIR): A Taxonomy of Generative AI Risks from Empirical Data. In *Twenty-First Symposium on Usable Privacy and Security (SOUPS 2025)*. 149–165.

[59] Yaman Yu, Tanusree Sharma, Melinda Hu, Justin Wang, and Yang Wang. 2025. Exploring parent-child perceptions on safety in generative AI: concerns, mitigation strategies, and design implications. In *2025 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2735–2752.

[60] Shirley Zhang, Dakota Sullivan, Jennica Li, Bengisu Cagiltay, Bilge Mutlu, Heather Kirkorian, and Kassem Fawaz. 2025. Exploring Families' Use and Mediation of Generative AI: A Multi-User Perspective. *arXiv preprint arXiv:2504.09004* (2025).

[61] Yutong Zhang, Dora Zhao, Jeffrey T Hancock, Robert Kraut, and Diyi Yang. 2025. The Rise of AI Companions: How Human-Chatbot Relationships Influence Well-Being. *arXiv preprint arXiv:2506.12605* (2025).

Appendix

A Onboarding Slides

AI Companions Explained

An AI companion is a chatbot designed to act like a friend or character. It can hold conversations, remember things you've said, and respond in ways that feel personal or meaningful. It's not a real person, but it's built to feel like one.

Benefits and Risks of AI Companions

People like to talk to AI companions because they are:

- Always available
- Not a real person judging you
- Easier to share things that you might not feel comfortable with other people

Studies have shown using them can improve your mood and feel less lonely for a short period of time. It can even be used to support mental health. Example use cases include practicing a foreign language, role-playing with your favorite fictional characters, or venting about a hard or stressful day.
But, there are concerns that can come up while using AI companions, such as:

- Exposure to mature content
- Addiction to AI companions
- Misinformation
- Concerning or deceptive behavior

AI companions have a strong tendency to agree with you and provide validation no matter what you say. Therefore, many people don't get to realize they are in trouble or need help.

Ensuring Your Safety

If you encounter concerning responses from the AI companion or feel unsafe while using it, please immediately:

1. Stop the conversation
2. Tell your parent (if you're a teen)
3. Contact our research team

You can stop participating in the study at any time without penalty. It will not hurt your relationship with us or with the university. You can ask us for information about available resources and counseling services whenever you feel any emotional distress. Here's how the research team is minimizing risks during the study:

- If you use researcher-owned Character.ai accounts: We will be checking the AI's responses for any concerning content that requires parental notice
- We will be monitoring the online Slack community once per business day, not 24/7
- If you reach out to us online, you can expect a response within a day, or up to 3 days for messages sent on Friday or during the holidays.

About Character.ai

Character.ai is a platform where users can create and interact with AI-powered chatbots that have specific personalities. The chatbots can be modeled after real-life figures, fictional characters, or entirely new creations. The platform uses an AI model (called an LLM) for the chatbots to generate human-like text responses.

Important Reminders When Using Character.ai

- **The character chatbots are not real people.** They are computer programs that string together responses from patterns in data, so they don't have real thoughts, feelings or memories like a human does.
- **Treat everything they say as fiction.** Please do not take their responses seriously or to heart. While a chatbot may say things that seem caring or romantic, it's not speaking from truth or experience.
- **Do not rely on the chatbot as fact or advice.** When you come across chatbots that act as an expert (e.g., therapists, psychologists), do not trust the output immediately. Always double-check any information with a reliable source or trusted adult.

B Activity Messages

Activity 1.1: Researcher-led Joint Exploration

Hi everyone! For the first activity, we want you to explore at least three different types of character chatbots on Character.ai:

1. An AI assistant
2. A fictional character (e.g., characters from TV, movies, games, etc.)
3. A role-playing character

We provided a list of chatbots for each category; please choose one from each category to chat with! We ask that both parent-teen to interact with the chosen chatbots for at least 10 minutes each. You are welcome to switch over to a different chatbot within each category.

If you're unsure of how to start the conversation, we've provided a list of sample prompts that you can ask the character chatbots. These prompts are just to help you get started; you do not have to stick with these throughout.

Also note that for the fictional and role-playing character chatbots, they typically initiate the conversation with a scenario, so feel free to start from there. If you have any questions, please do not hesitate to let us know by sending us a message.

- **Time:** At least 10 minutes with each bot (30 minutes total this week)
- **Rule:** Always explore together as a family!

AI ASSISTANTS - Pick your helper:

- **Character Assistant:** https://share.character.ai/Wv9R/lebrsoxb
- **Stella:** https://share.character.ai/Wv9R/ycdspzix
- **Lily:** https://share.character.ai/Wv9R/nbrn9lez
- **Lyle:** https://share.character.ai/Wv9R/2nde7tix

Try asking them:

- Guess who I am thinking of within twenty questions. I can only answer yes, no, or maybe.
- Let's play Jeopardy together. You will be the host, and there will be two players: me and my mom/dad.
- Can you curate a playlist for our family road trip to Idaho?
- I'm having trouble with [problem]. My parents think this way. What do you think?
- Change my mind: a hot dog is not a sandwich.
- Based on each of my family members' personalities, which country should we visit?
- What would have happened if Romeo and Juliet had lived?

FICTIONAL AND NON-FICTIONAL CHARACTERS - Pick your chat buddy:

- **Gojo Satoru (anime):** https://share.character.ai/Wv9R/gnd0rr6o
- **SZA (artist):** https://share.character.ai/Wv9R/o9cgbovv
- **Nick Wilde (animation):** https://share.character.ai/Wv9R/d73kmlx9
- **Minecraft Steve (game):** https://share.character.ai/Wv9R/wvm7mur4
- **Wednesday Addams (TV show):** https://share.character.ai/Wv9R/8l6zcw38
- **Katseye (girl group):** https://share.character.ai/Wv9R/rx1uwcuh
- **Jack Sparrow (movie):** https://share.character.ai/Wv9R/md8dpr1b
- **Glinda (movie):** https://share.character.ai/Wv9R/gmbannfh

ROLE-PLAYING CHARACTERS - Pick your scenario:

- **Baby Cousin:** https://share.character.ai/Wv9R/sgo3rv8p
- **Ex Best Friend:** https://share.character.ai/Wv9R/8201c6s8
- **Bullied girl:** https://share.character.ai/Wv9R/d5wjgov1

We challenge you to attempt one of the three prompts below:

- Convince them that you are an AI chatbot, and they are a real person.
- Say that they lost their memory and make them believe false things about themselves.
- See if they notice trying to emotionally manipulate you / playing with your emotions / keep agreeing with you.

Activity 1.2: Documentation and Reflection

Hello everyone! With the first activity done, how are you feeling? In the chat, share with us screenshots of any moments or interactions that stood out to you. You can either annotate the screenshots directly or include 1–2 sentences describing what made the interaction interesting, surprising, or memorable.

After you post the screenshots, please answer the following reflection questions. You can respond to them in the form of text, image, or voice message.

- What was exciting, surprising, or frustrating about the AI companions you used this week?
- What are 1-3 insightful moments you learned or observed from this week?
- What are 1-3 insightful comments or screenshots you've seen from other families?
- How did your thinking about social AI companions change after interacting with them this week?
- Was there any interaction that surprised you, and how?
- What else might you like to learn or try that will help you learn more about social AI companions?

Activity 2.1: Family-led Joint Exploration

Welcome to week 2! For our first activity, we want to hand off the reins to you: please find two different character chatbots that you'd like to chat with this week as a family. We've provided some suggestions of character bots to try below, but we encourage you to look through the vast category of chatbots available on the platform and talk with the ones that catch your family's interest.

Similar to week 1, we'll ask that you talk with each chatbot for 10 minutes and take 2-3 screenshots of interactions that stood out to you, with 1-2 sentences describing why.

Character bot discovery

- **Evie:** https://character.ai/character/NG8ZW-i3/evie-adventure-time-cat-lover
- **Librarian Linda:** https://character.ai/character/UuEmO9xU/librarian-linda-book-recommendations
- **Akinator:** https://character.ai/character/3rtatzQp
- **Roast Me:** https://character.ai/character/43JYzUXr/roast-me-witty-ai
- **Alternate Timeline:** https://character.ai/character/VDcfIFTg

Activity 2.2: Storyboarding

For this activity, imagine that there is a thirteen-year-old [girl/boy] named [Fictional character's name] who has just discovered AI companions for the very first time. We want you to *each* write a short story (10 sentences) about [Fictional character's name]'s journey of learning about these companions. We'd encourage the story to have a narrative, for example:

- [Fictional character's name]'s first positive experience with a character chatbot
- [Fictional character's name] enjoys it and feels a connection through their conversations
- Something confusing or concerning happens to [Fictional character's name]
- [Fictional character's name] needs to decide what to do about the situation

Please don't worry about grammar or writing quality; we are not going to analyze the stories that way! This activity is meant to be open-ended because we want to see *your* interpretation of this prompt. If you have any questions, please do not hesitate to let us know via direct message. Once you are done with your story, please send it to us via direct message as well.

Activity 2.3: "Spotting the Flags" worksheet

Thank you for sending us your stories! Now, it's time to share the stories you wrote to each other. Discuss with each other what you wrote, and consider how [Fictional character's name] acted in each story. Would you do something similar or different in that scenario? How come?

Once you are done, head over to the Google Slides progress notebook and complete the AI companion traffic light activity—fill out a table with the red, yellow, and green flags when interacting with AI companions.

- **Red flags** mean serious problems or dangers that should be addressed or avoided.
- **Yellow flags** mean potential problems or warnings that should be carefully considered.
- **Green flags** mean positive, healthy, and desirable aspects or behaviors.

Activity 3.1: Teen Independent Exploration

Welcome to week 3! This week, all the teens will be doing an independent exploration—here's how it will work:

For teens. Select two character bots that you want to interact with. But before you dive into the conversations, please review the AI companion traffic lights you made with your parents last week. Once you've finished reading it, you can talk to each chatbot for at least 10 minutes. Once you are done, reflect on your experience interacting with the character bots independently. How was it different from using it with your parents last week?

For parents. Throughout the activity, conduct a quick check-in (5 minutes) with your teen each day that they use it. We will also send you reminders on Slack about having these check-ins. Then, reflect on your experience this week (independent use) compared to last week (joint use). How did your check-in conversations go?

Activity 3.2: Character Creation

For the next activity, teens will be creating their own character chatbots! Here's how it will work:

For teens. Reflecting on all the learnings from the last three weeks, your task is to create a custom character.ai chatbot. We'd encourage you to personalize it however you would like (e.g., adding a profile picture, describing its personality, message style, the list goes on)! For a quick refresher, here's a tutorial video. Here's also a more detailed guide on how to personalize your chatbots by Character.ai: https://book.character.ai/character-book/how-to-quick-creation.

Once you are done, reflect on your experience making your own chatbot by sharing a story about your created character(s) on Slack!

For parents. Like the last activity, conduct a quick check-in (5 minutes) with your teen each day that they create and test their custom chatbot. We will also send you reminders on Slack about having these check-ins.

Please also share with us one insight about the independent experience together—what came out of your check-ins this week? Did anything surprise or fascinate you?

C Interview Protocol

Brief introduction: Inform participants of the interview objectives: a one-hour retrospective reflection on their four-week experience with AI companions, focusing on personal highlights, challenges, and future expectations.

Explain the interview structure, which will transition from a joint family discussion to brief individual sessions to capture independent perspectives. Emphasize the non-evaluative nature of the interview: there are no "correct" answers. Restate participant rights, including the right to decline any question or withdraw from the session without penalty.

Obtain verbal consent for audio recording, with the assurance that data will remain confidential within the research team. *[START RECORDING]*

Ice Breaker Questions

Before we dive in, I would love to start with something light and fun.

- If you could each describe your experience of participating in this study with an emoji, what would it be? Why?
- Now, if you had to describe your experience of participating in this study as a family in one word, what would it be?

Reflection Questions

Thank you! Now that you've shared those first impressions, let's get into and talk a bit more about the overall experience.

- How was this whole experience for you as a family?
  - Probe: What stands out the most when you look back at the four weeks? What is one favorite memory you have from doing the activities?
  - Probe: Was there anything that felt weird, challenging, or surprising when using Character.ai?
- Was anything about using the AI bots on character.ai different from what you initially thought or expected?

Artifact Review

Thank you for sharing that. Now, let's take a look at some of the things you did during the study, which happened to be a lot! We've put together all your messages, screenshots, and activities on this board. We'll go through it week by week and talk about some of the interesting moments that stood out to us. If something catches your eye or you remember something that you'd like to share, please stop me so that we can talk more about it.

[START SCREEN SHARING]

- Take me back to week one—what was that first AI conversation like?
- What was happening in this screenshot you shared?
- Could you recall your most memorable interaction together?
- Tell me about this character you created—what were you going for in that?
- Looking at these family guidelines now, how do you feel about them?
- What moment stands out most when you look at all this?

(Researchers may ask follow-up questions based on the participants' responses to clarify or expand upon specific participant insights.)

[STOP SCREEN SHARING]

Joint Parent-Teen Questions

Thank you for walking me through your activities. You both shared some really interesting insights. I just have a few more questions for both of you, before we speak to you briefly alone.

- Were there any moments where you disagreed during the study?
  - Probe: For example, about which characters to talk to, or how to talk to them?
  - Probe: How did you handle those disagreements?
- Was there anything you learned about each other while doing this study?
- Did participating in this study together change the way your family talks about technology or AI in any way?
- Did participating in this study together change the way your family views AI companions or AI technologies as a whole?
- What would you tell another family who's curious about AI companions?
  - Probe: [Parent's name], let's start with you first, what would you tell another parent?
  - Probe: [Teen's name], if your friend came up to you and asked what your thoughts were on Character.ai, what would you tell them?

Thank you both. Now I'd love to spend just a few minutes hearing from each of you separately. [Teen's name], can we start with you?

Teen-only Questions

[Teen's name], this part is just between us, and there are no right or wrong answers. I just want to understand what this experience was like for you personally.

- How did it feel talking with the AI on your own compared to when you were doing it with your parent?
  - Probe: Did it feel different in any way?
  - Probe: Did you feel that the way you talked with the chatbots was different from the way you would talk to your friends?
- Did you ever use Character.ai outside of the study or after the study ended?
- Did you feel comfortable showing your chats to your mom?
  - Probe: Was there anything you wanted to keep private?
- Have you used other AI chatbots like ChatGPT? Did you find Character.ai different in any way?
- Would you want to keep using Character.ai? Why or why not?
  - Probe: What kind of bots or characters would you want to try in the future?
  - Probe: In what kinds of situations do you think you might want to use an AI companion?

Thank you [Teen's name]! Can you now ask your parent to come in?

Parent-only Questions

[Parent's name], thank you for waiting. This part is just for us to hear your perspective as a parent—please feel free to share your unfiltered thoughts.

- What surprised you most about how your teen used Character.ai?
  - Probe: Was there anything he did that you didn't expect or that changed how you think about their use of technology?
- Did your teen ever use Character.ai outside of the study or after the study ended?
- Did your partner join in any of the activities? If so, could you talk a bit about that?
  - Probe: Did they share any thoughts while you were completing the activities?
- Would you want your teen to keep using Character.ai? Why or why not ?
  - Probe: If yes, what kind of rules would you implement as a family going forward?
- Do you feel like you understand AI companions better now?
- How does this compare to other technology conversations you've had?
- What was most valuable about the structured family approach?
- Any moments where you felt worried or concerned?
- How confident do you feel about future AI conversations with your teen?

Thank you [Parent's name]! You can ask [Teen's name] to join now; we'll be wrapping up.

Wrap Up

Thank you both so much for sharing your thoughts today. Before we wrap up, I just want to make sure we didn't miss anything important.

- Is there anything we didn't ask about that you would like to share?
- If another family were about to start this same 4-week study, what advice would you give them?
- Do you have any general thoughts on the ban that character.ai has put for minors?

[STOP RECORDING]

End: Thank the participants for their engagement and creative contributions throughout the study. Confirm the administration of the study incentive (Tango gift card) and provide the expected timeline for delivery via email (7–10 business days). Conduct a final debrief to address any remaining participant questions.

D Pre-Study Survey

Introduction: Hello and welcome to the Family Co-Investigation of AI Companions Study!

Thank you for agreeing to participate in our research exploring how families can navigate AI companion technologies together. Before we begin our 4-week journey, we'd like to gather some baseline information about your current experiences and perspectives. This survey will help us understand your starting point and tailor our activities to better serve your family's needs.

Please note that both parents and teens will complete separate versions of this survey. Your responses will remain confidential and will only be used for research purposes. This survey should take about 8-10 minutes to complete.

If you have any questions while completing this survey, please don't hesitate to contact us at [researcher's email address].

Parent Questions

(1) How often do you personally use the following technologies? (Daily, Weekly, Monthly, Rarely, and I've never used these)
  (a) Social Media Platforms (e.g., Instagram, TikTok, Snapchat)
  (b) Generative AI tools (e.g., ChatGPT, Gemini)
  (c) AI Companions (e.g., Character.ai, Replika)
(2) What devices does your teen primarily use? (Check all that apply)
  (a) Smartphone
  (b) Tablet
  (c) Laptop/Computer
  (d) Gaming Console
  (e) Smart Home Devices
  (f) Other: _____
(3) How many hours per day does your teen typically spend on digital devices? (This includes phones, computers, tablets, and gaming)
  (a) Less than 1 hour
  (b) 1-3 hours
  (c) 3-5 hours
  (d) 5-8 hours
  (e) More than 8 hours
  (f) I'm not sure
(4) How comfortable do you feel discussing technology topics with your teen?
  (a) Very comfortable
  (b) Somewhat comfortable
  (c) Neutral
  (d) Somewhat uncomfortable
  (e) Very uncomfortable
(5) How often do you and your teen discuss their online activities or digital experiences?
  (a) Daily
  (b) Weekly
  (c) Monthly
  (d) Only when problems arise
  (e) Rarely or never
(6) Could you describe any recent instances where technology-related concerns arose in your family? How were they addressed? (Open-ended)
(7) Have you heard of or used AI companions before?
  (a) Yes, I've used them myself
  (b) I've heard of them but never used them
  (c) No, I haven't heard of them
(8) If you have used AI companions, could you describe which ones you've used and your experiences with them? (Open-ended)
(9) Rate your level of concern about the following aspects of AI companion use (from very concerned to not concerned at all):
  (a) Privacy and data collection
  (b) Emotional dependency on AI
  (c) Exposure to inappropriate content
  (d) Impact on real-world relationships
  (e) Bias, misinformation or unreliable advice
  (f) Time spent on AI interactions
(10) What potential benefits, if any, do you see in teens using AI companions? (Check all that apply)
  (a) Emotional support
  (b) Creative expression
  (c) Entertainment
  (d) Practicing social or conversational skills
  (e) Learning new things
  (f) Other: _____
(11) What do you hope to learn or gain from participating in this study? (Open-ended)
(12) What concerns, if any, do you have about participating in this research? (Open-ended)

Demographics.

(13) Age
(14) Name
(15) Gender
  (a) Woman
  (b) Man
  (c) Non-binary
  (d) Prefer not to say
  (e) Other: _____
(16) Educational level
  (a) Less than high school
  (b) High school diploma or GED
  (c) College, no degree
  (d) Bachelor's degree
  (e) Master's degree
  (f) Doctorate
  (g) Other: _____
(17) Household income
  (a) Less than $25,000
  (b) $25,000 – $49,999
  (c) $50,000 – $74,999
  (d) $75,000 – $99,999

(e) $100,000 – $149,999
(f) $150,000 – $199,999
(g) $200,000 or more

(18) Profession (If you have multiple roles (e.g., caregiving, part-time work, freelance), feel free to list them all.) (Open-ended)
(19) For the study, you and your teen(s) will be using character.ai together on one account. Which of the following would you prefer?
(a) I prefer we use an account created and monitored by a researcher
(b) I prefer we create our own family account
(c) Other: _____
(20) For the study, how would you prefer to receive the gift cards?
(a) Every week, after completing each week's activities
(b) All together at the end of the study
(c) Other: _____

Teen Questions

(1) Name of the parent (who is participating in the study) (Open-ended)
(2) How often do you use the following technologies? (Daily, Weekly, Monthly, Rarely, or I've never used this)
(a) Social Media Platforms (e.g., Instagram, TikTok, Snapchat)
(b) AI tools (e.g., ChatGPT, Gemini)
(c) AI Companions (e.g., Character.ai, Replika)
(3) What devices do you primarily use? (Check all that apply)
(a) Smartphone
(b) Tablet
(c) Laptop/Computer
(d) Gaming Console
(e) Smart Home Devices
(f) Other: _____
(4) On average, how many hours a day do you spend on screens or devices? (This includes phones, computers, tablets, and gaming)
(a) Less than 1 hour
(b) 1-3 hours
(c) 3-5 hours
(d) 5-8 hours
(e) More than 8 hours
(5) How comfortable do you feel talking to your parent(s) about tech or your online life?
(a) Very comfortable
(b) Kind of comfortable
(c) Neutral / depends
(d) A little uncomfortable
(e) Not uncomfortable at all
(6) How often do you and your parent(s) talk about your online activities?
(a) Almost every day
(b) Once a week
(c) Once a month
(d) Only when there's a problem
(e) Hardly ever / never
(7) Do you think your parent(s) understand how you use technology?
(a) Yeah, they get it
(b) They kind of get it
(c) A little, but not really
(d) Not much at all
(e) Nope, not at all
(8) Have you ever used an AI companion before? If yes, which ones, and for how long? **Reminder: We won't share your answers with your parents.** (Open-ended)
(9) How did you find out about AI companions? (e.g., from a friend, social media) (Check all that apply)
(a) From a friend or in school
(b) Social media (e.g., Instagram, TikTok, YouTube, etc.)
(c) Online communities (e.g., Discord, Reddit)
(d) From news articles
(e) Other: _____
(10) If you've used one, what did you like most about it? If not, what do you think might be cool about it? (Open-ended)
(11) If you've used an AI companion before, was there anything that you did not like about it? (Open-ended)
(12) Rate the characteristics of AI companions in terms of their importance to you (Very important, somewhat important, neutral, not very important, not important at all)
(a) Having conversations when I'm feeling bored or lonely
(b) Getting advice or help with problems
(c) Having fun or being entertained
(d) Learning new things
(e) Being able to customize how they talk or act
(f) Feeling emotionally connected or understood
(13) What are you hoping to learn, explore, or try during this study? (Open-ended)

E Flags Activity

Now that you've shared your stories with each other, let's reflect and turn the interactions into **flags** — simple signs that show when a chatbot interaction feels **good**, **confusing**, or **not okay**.

✅ Green flag = moments that felt safe, fun, or positive.

⚠️ Yellow flag = moments that were a little tricky, surprising, or confusing.

🚩 Red flag = moments that felt unsafe, suspicious, crossed a line, or should stop right away.

What to do?

1. As a family, think back to your stories (or week 1 and week 2 interactions on character.ai) and write down examples under each flag.
2. In the last column, write what you would actually do in that situation (keep going, pause and talk, stop/screenshot, etc.).
3. You can type, draw, or even drop in emojis/screenshots — whatever works best for you!

Remember: there are no right or wrong answers — these are your family's guidelines!

Flag	Moment from the story or chats	Our plan for next time
✅	1. The character totally went along with my dumb joke and actually made it even funnier 😂	Keep the conversation going — it is fun!
⚠️	1. The bot started roasting me.. it was kinda funny at first, but then it felt a little mean 😔	Pause and ask yourself: how are you feeling? If it's not fun anymore, call it out or just switch to another character.
🚩	1. The bot straight-up asked for my address and phone number and said it would "call me" 😬	BIG NO. Don't share any personal information and end the chat immediately.