ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events

Woosuk Seo*
University of Michigan
Ann Arbor, USA
seow@umich.edu

Chanmo Yang
Wonkwang Univ. Hospital &
Wonkwang University
Republic of Korea
ychanmo@wku.ac.kr

Young-Ho Kim
NAVER AI Lab
Republic of Korea
yghokim@younghokim.net

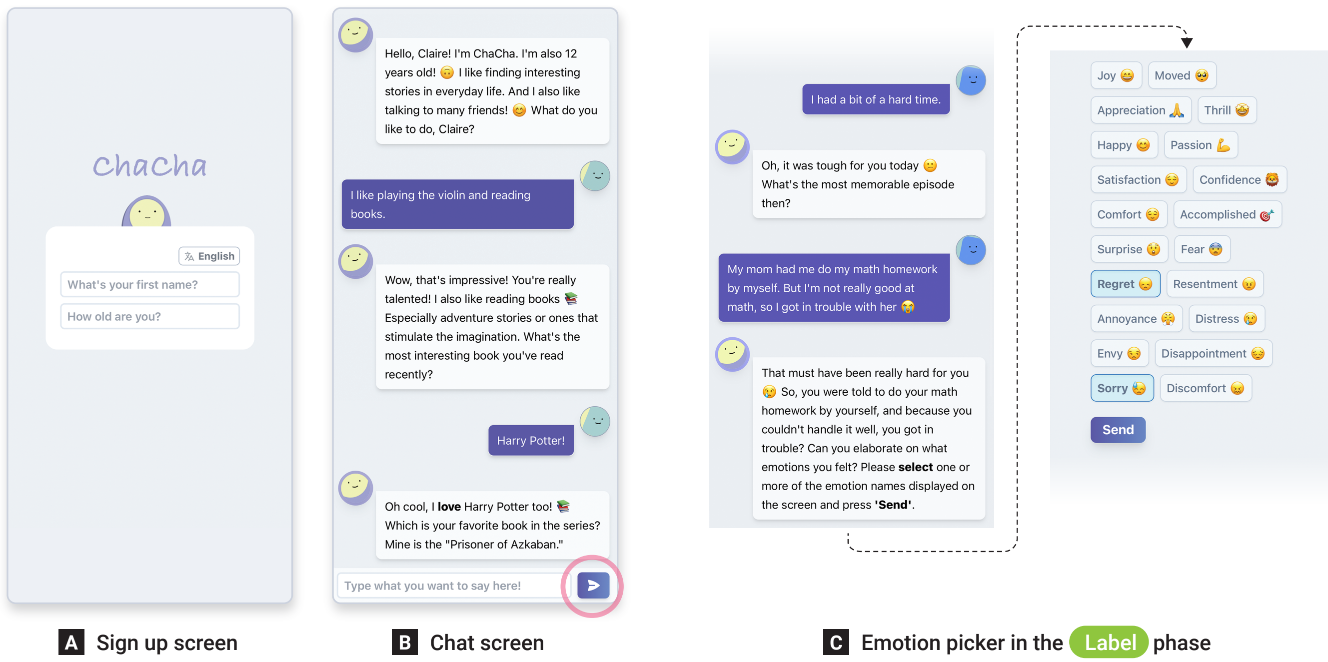

Figure 1: Main screens of ChaCha interface. A The user can initiate a new session by entering their name and age, and B converse with ChaCha. C If the user struggles with describing their emotions, ChaCha offers a list of emotions for guidance.

ABSTRACT

Children typically learn to identify and express their emotions by sharing stories and feelings with others, particularly family members. However, it is challenging for parents or siblings to have effective emotion communication with children since children are still developing their communication skills. We present ChaCha, a chatbot that encourages and guides children to share personal events and associated emotions. ChaCha combines a state machine and large language models (LLMs) to keep the dialogue on track while carrying on free-form conversations. Through an exploratory study with 20 children (aged 8–12), we examine how ChaCha prompts children to share personal events and guides them to describe associated emotions. Participants perceived ChaCha as a close friend and shared their stories on various topics, such as family trips and personal achievements. Based on the findings, we discuss opportunities for leveraging LLMs to design child-friendly chatbots to support children in sharing emotions.

*Woosuk Seo conducted this work as a research intern at NAVER AI Lab.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05…$15.00
https://doi.org/10.1145/3613904.3642152

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Empirical studies in HCI**.

KEYWORDS

Chatbots, Children, Large Language Models, Conversational Agents

ACM Reference Format:
Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA*. ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3613904.3642152

1 INTRODUCTION

In the context of child development, the perception and expression of emotions during early childhood hold significant importance. As children progress through different developmental stages, they become better at self-reporting their emotional experiences; they gradually develop the ability to express their emotions or withhold emotional expression to avoid potentially adverse reactions from others [72, 87]. Hence, children need developmentally appropriate education and practice to develop such emotional competencies. Children's relationships with their parents or siblings significantly impact their emotional development in the family context. Better quality of sibling relationship is more likely to promote children's higher social-emotional well-being [86]. Similarly, parents play an essential role in supporting how children identify and express their emotions [10, 37]. For instance, mothers' behaviors in providing emotional support or controlling their children's emotional states significantly impacted how children manage their anxiety [10].

Despite its significance, emotion communication [1] has not been frequently addressed in parenting interventions [75]. For parents, it is challenging to guide emotion communication with children who are still developing communication skills [75]. With the lack of support and interventions, parents would provide less emotional support, which may result in adverse mental health outcomes for children, such as anxiety [10]. Notably, only-child children may have more challenges than those with siblings since they would have fewer opportunities to build supportive sibling relationships and promote their emotional well-being (*e.g.*, [86]). Even worse, the COVID-19 pandemic has impacted children's social activities and daily routines, affecting their socio-emotional development [23, 36]: A recent survey study showed that, along with anxiety and emotional problems, children and adolescents (ages 8–17) have experienced more difficulty in emotional regulation during the pandemic [36].

In the HCI community, prior studies explored how technology can be designed to support children with learning and sharing their emotions [41, 73] or to enhance parents' awareness of children's emotional well-being [65]. These studies showed how a chatbot [73] or a tangible artifact [65] can be used to detect children's emotions more accurately. For instance, Santos *et al.* [73] demonstrated how a chatbot leverages a storytelling strategy to effectively recognize children's emotions from the given list of emotions. Although prior studies have presented opportunities for technology to promote children's emotional regulation through detection, they primarily focused on collecting data from children and sharing it with their parents, limiting our understanding of how technology can better support children in identifying and expressing their emotions. The existing technology was mostly limited to probing questions to recognize children's emotions rather than supporting them with practicing to express their emotions. Children's perceptions and preferences for how, when, and what to communicate about their emotions were often overlooked.

To explore how chatbots could support children in identifying and expressing their emotions, we present ChaCha (Figure 1), a chatbot that guides children in sharing their emotions about personal events. Following the emotion coaching guidelines [31], ChaCha elicits a child's key events, helps them label associated emotions, explores possible solutions to alleviate adverse emotional experiences, and encourages sharing their emotions with parents. For better engagement with children in the conversation, ChaCha's dialogue system leverages large language models (LLMs) [8, 61] to coherently follow up the conversation context while generating *in-situ* responses on serendipitous topics. Informed by formative interviews ($N = 6$) with child mental health professionals (*e.g.*, psychiatrists, psychotherapists), we designed the conversations to consist of multiple phases with sub-goals and a dedicated LLM to keep the LLM-driven conversation on track and easy for children to follow. With this LLM-driven chatbot, we aim to answer the following research question: *How feasible is an LLM-driven chatbot in prompting children to share their emotions about personal events?*

To examine how children interact with ChaCha and perceive it for sharing their emotions about personal events, we conducted an exploratory user study with 20 children in South Korea. During the study session, participants freely conversed with ChaCha on a smartphone for about 30 minutes. We observed that participants felt comfortable conversing with ChaCha, sharing diverse events ranging from a family trip to a conflict with their parents. With ChaCha's assistance, participants were able to recall and identify their emotions associated with the positive and negative events. Participants also recognized ChaCha was skilled in listening to their stories about events and empathizing with their emotions, hence perceived ChaCha as a close friend with whom they would share emotions that they would not share with others.

In summary, this work offers the following contributions:

(1) The design and implementation of ChaCha, a novel LLM-driven chatbot that helps children share their emotions about events through free-form conversations. ChaCha's design was informed by formative interviews ($N = 6$) with children's mental health professionals. We present design rationales of a chatbot for encouraging children to express their emotions and conversational flow of ChaCha to meet those rationales. The source code of ChaCha is available at `https://naver-ai.github.io/chacha`.

(2) An empirical understanding of how children recognize and interact with ChaCha. We conducted an exploratory study with 20 children. From the analysis of the dialogue logs and interview transcripts, we provide findings on how children interact with a chatbot driven by LLMs in sharing emotions about personal events.

(3) Design considerations for LLM-driven chatbots for children. We present the benefits of leveraging LLMs in designing child-friendly chatbots, such as showing empathetic behaviors that encourage children to share their emotions, and suggestions to mitigate potential concerns about using LLMs, such as balancing chatbots' self-disclosure and overreliance on chatbots.

2 RELATED WORK

In this section, we explore prior studies about supporting children's emotional development and communication and examine existing technology for children's emotion regulation and mental health. We

[1] Hereinafter, we define emotion communication as the process of recognizing and sharing emotions between two or more individuals.

then review previous work about the characteristics of LLM-driven chatbots, presenting the potential advantages of LLMs in children's mental health.

2.1 Emotion Communication and Coaching

The abilities to identify and express emotions are two essential skills for emotional competence discussed in the field of children's emotional development [72]. From infants, children begin to learn expressive behaviors that are emotion-related components [71]. Along with cognitive and social skills, emotional competence can be developed with proper support and age-appropriate education as children gradually acquire more knowledge and experiences. Due to this nature of emotional development, parents' emotion communication with children essentially affects the development of children's skills to recognize, express, and share their emotions. Ample literature in psychology and medicine has emphasized how children's emotional development can be affected by how parents communicate about children's emotions [4, 10, 18, 37]. In particular, middle childhood (ages 6–12) is considered a necessary time to foster open communication with parents. Children in this age range are old enough to decide how to express which emotions to with others [70], while they still need to practice and develop skills to express their opinions accurately [67]. In addition, middle childhood is when children spend more time away from their parents than younger children. As a result, parents would have less awareness of their children's negative experiences unless children choose to discuss them [26]. Parents may face considerable challenges in providing adequate support to their children when they lack an understanding of children's feelings [74].

To support parents in facilitating emotion communication and regulation of children, a renowned psychologist, John Gottman, developed a guideline for emotion coaching to help parents teach their children how to identify and express their emotions [31]. The guideline consists of five steps: (1) Be aware of the child's emotions, (2) Recognize the child's expression of emotions, (3) Listen with empathy, (4) Help the child learn to label their emotions, and (5) Set limits when helping the child to solve problems [31]. Although emotion coaching is a widely used and validated framework [33], parents often face various challenges in thoroughly learning and applying the principles of emotion coaching to their contexts [35]. Such challenges may include a lack of experience with emotion communication in their own childhood, a lack of verbal skills to name the emotions that their children express, and fear of letting their children verbalize specific emotion words [35]. To complement parents' efforts in emotion communication with children, prior studies have proposed interventions for children to support them in sharing their experiences and emotions [34, 35].

2.2 Chatbots for Children's Emotion Regulation and Mental Health

Technological interventions have supported children with emotion regulation and mental health (*e.g.*, mHealth apps [58, 59]). Among various types of technology, chatbots have lately shown promising opportunities in promoting emotion regulation and mental health of children and adolescents [20, 73, 78]. Through conversational interactions, chatbots offer a space where children can freely share their stories and express their emotions in their own words while keeping their secrets from others [47].

Acknowledging the benefits of conversational interactions, prior work has proposed chatbots as a promising way to assist children's emotion regulation or promote their mental health. Dosovitsky *et al.* [20] developed BethBot, a chatbot that includes cognitive behavior therapy (CBT) intervention modules used in psychoeducation to educate adolescents on depression and teach coping skills. Based on the analysis of user experience questions, Dosovitsky *et al.* [20] found that adolescents perceived BethBot as an acceptable mental health resource that would improve their symptoms. Santos *et al.* [73] developed a chatbot by adapting a storytelling strategy. The chatbot recognizes specific keywords from children's responses to determine their emotions on the given list of emotions. The user testing with children (ages 9-11) revealed that the child participants felt comfortable sharing their stories with the chatbot and showed how accurately the chatbot could detect children's emotions based on their storytelling. Although prior work has posed the benefits of leveraging chatbots for children, existing chatbot research in this domain predominantly adopts a rule-based approach [1] where the chatbot follows a predefined conversational flow with close-ended questions and templated messages. Hence, the chatbot may not be responsive to children's messages that are beyond the scope of the designer's prior considerations [1, 38]. Given the open-ended nature of sharing emotions and personal events, we believe that free-form conversations could be more effective for carrying on communication about children's emotions. Thus, our study extends the line of research about chatbots for children's mental health by demonstrating the opportunity for LLMs to enhance the open-ended conversation capability of a chatbot.

2.3 Large Language Models for Conversation

The recent advance of pre-trained LLMs (*e.g.*, GPT [8, 61], PaLM [15], LLaMA [77], LaMDA [76], HyperCLOVA [46]) has presented new opportunities for designing and improving various natural language processing tasks [55, 88]. Specifically, LLMs led to the invention of a new design approach for chatbots, complementing the limitations of traditional rule-based and retrieval-based approaches [38, 45, 80]. When used to bootstrap a chatbot, LLMs generate a message that naturally continues the current dialogue by augmenting the content generation process with their large-scale pre-trained knowledge. Hence, the LLM-generated messages tend to be coherent across the conversation session and human-like, which is more flexible than rule-based chatbots. Unlike retrieval-based chatbots that require a number of domain-specific dialogue data to train a retriever [38], only a few examples or even no examples can bootstrap comparably working chatbots [80]. For instance, Wei *et al.* presented GPT-3-driven chatbots devised with simple zero-shot prompts to collect health-related questionnaires through free-form conversation [80]. More recently, enterprises introduced LLM-driven chatbot services (*e.g.*, ChatGPT [60], Bard [30], HyperClovaX [6], Pi [40]) that enable individuals to perform a variety of tasks, such as idea generation and data processing, via conversational interactions [21].

Adopting this benefit of leveraging LLMs, recent studies have explored the potential use of LLMs in mental health contexts (*e.g.*, ChatGPT-generated patient-psychiatrist simulations [13]). LLMs may also express personality traits [44], maximizing their content generation capability for a specific context and target. This aspect of LLMs could be helpful in facilitating a free-form conversation about emotional distress or psychological issues by simulating a more sociable, attentive, or caring personality. However, further studies are required to examine the feasibility of this relatively new technology in promoting mental health, particularly for children. A recent study presented how an LLM-driven agent can help children generate curious questions, potentially suitable for training their question-asking skills [2]. Similar to this case of using LLM for skill development, LLM-driven conversations could provide a space for children to communicate about their emotions more openly rather than simply answering questions by rule-based chatbots. Such open communication in free-form conversation could help children develop their skills to identify and express their emotions in their own words. Motivated by such potential of LLMs, we examine how an LLM-driven chatbot can prompt children to share their emotions about specific events they experience.

3 FORMATIVE INTERVIEWS

To understand clinical practices for communicating emotions with children, we conducted semi-structured interviews with six child mental health professionals based in South Korea. The interviewees were three child psychiatrists, two child psychotherapists, and one child development specialist. We recruited them from university hospitals, child psychotherapy centers, and a private child mental health clinic. The interviews lasted about 1 to 1.5 hours, and two researchers joined the interviews. The interviews were conducted in person or remotely based on each professional's availability. We offered 100,000 KRW (approx. 80 USD) as compensation.

We asked about the professionals' perspectives on the challenges children face when sharing their emotions, strategies to better communicate with them about their feelings, and a chatbot that supports children in sharing emotions. As a probe, we showed the professionals a video prototype of a conversation between a chatbot and a child (Figure 2a; See supplementary material). The purpose of the video was to help interviewees understand the expected conversations between children and a chatbot. We also asked the interviewees to provide any materials or tools they refer to or use when communicating with children (See Figure 2b). The interviews were audio recorded and transcribed. The first author open-coded transcripts, going through a few iterations. The full research team then iterated multiple sessions of discussions identifying themes regarding design considerations. Through this process, we finalized three design considerations based on professionals' approach to emotion communication with children.

We identified three considerations for designing a chatbot for emotion communication with children: First, **a chatbot should show its empathy with children's emotions.** The professionals articulated the importance of empathy in helping children define and express their feelings. A common strategy to empathize with children's feelings is to *repeat what they have shared* (*e.g.*, "*I see. You were sad when you lost the game*"). Specifically, it is important to empathize with children from their level of understanding, focusing on what children actually feel rather than judging their feelings. All of the professionals mentioned that it seems to be appropriate to have a peer persona for a chatbot. One of them particularly noted that a peer chatbot would allow children to open themselves and share their emotions more easily. Also, children would be more accepting of the suggestions that the peer chatbot provides to them. Second, **a chatbot should give children options for emotions to choose from.** Some children, particularly younger ones, answer questions with just one word (*e.g.*, yes/no, good/bad). To probe more questions in such cases, some professionals shared that they often give potential options. Adopting this approach, chatbots could help children who express difficulty in describing their emotions by providing some options for emotional words that they can choose from. Some professionals also noted that these options for emotional words can be helpful to children to explore various emotions. Lastly, **a chatbot should encourage children to share their feelings with their parents.** The professionals envisioned the ideal role of

(a) Interview settings **(b) Example emotion cards**

Figure 2: Photos taken in one interviewee's office: (a) The interview settings with a book that the interviewee showed and a tablet of the experimenter to demonstrate a video prototype, and (b) emotion cards that the interviewee uses when guiding children to describe their emotions. The emotion keywords on the envelopes are anger, anticipation, joy, sadness, fear, and surprise (from left to right).

the chatbot should be a bridge or a facilitator between children and their parents, not replacing the parents' role in children's emotional development. Hence, the chatbot should not simply be the one who solves and teaches everything that children ask. It should guide the children to share their emotional needs with their parents or healthcare providers so that they can receive relevant support.

Along with the three design considerations described above, the findings from the interviews with the professionals informed us of the conversation flow of the chatbot and the initial ideas of the chatbot's functions, which we will describe in Section 4. To carefully design the conversational flows and the chatbot's behaviors, we invited one of the psychiatrists we interviewed as an author of this work and collaborated with him on the design of the chatbot and the study protocol.

4 CHACHA

Informed by the formative interviews, we designed and developed ChaCha, a chatbot to prompt children to share their stories about specific events and associated emotions. We improved the interface and the underlying dialogue system by running pilot sessions with two children recruited by word of mouth. This section describes the design rationales for ChaCha and the dialogue system, along with implementation details.

4.1 Conversation Design Rationales

DR1. Empathize children from their perspectives. Our target audience is child users who experience the world differently from adults rather than lack competencies [9]. Especially, children (age 6-12) begin to develop competencies and preferences in sharing emotions [70]. As one of such preferences, children prefer to share emotions, which tend to change from adults such as parents to their peers [69]. For instance, in a sample of children (age 6-14), older children would share emotions with their peers rather than with their parents, whereas younger children preferred adults to peers [69]. Aligning with the findings from prior work, the insights from our interviews with professionals also advocate for the use of a peer persona for the chatbot. Thus, we decided to provide children with a peer interaction experience. By setting ChaCha as a peer child, we expected the users to have more comfortable conversations about their stories and emotions. Moreover, based on this child persona, we tried to control the range of information that ChaCha shares with children. For instance, we gave specific instructions to ChaCha to only answer the questions that are somewhat related to children's interests or their stories and direct them to ask their parents if they ask inappropriate questions (*c.f.*, Appendix B). We leveraged an LLM to determine the boundary of child-appropriateness. For safety considerations, we conducted internal tests on ChaCha's responses to questions necessitating parental guidance (e.g., regarding violence) prior to initiating the user study.

DR2. Support children explore a range of emotions. From the formative interviews, the professionals articulated the importance of allowing children to explore a range of different emotions. More importantly, negative emotions may need to be probed further since those emotions usually involve various stakeholders and factors (*e.g.*, other friends, recurring occasions). It is essential to explore negative emotional experiences more carefully than positive ones so that children can learn how to manage negative emotions or ask for help. Through sharing the event that caused the negative emotions, children may develop different strategies to change their behaviors and goals that ultimately help them manage and alleviate their negative emotions. [17]. Thus, we branch out two phases to handle each category of emotions: Find for negative emotions and Record for positive emotions. More detailed descriptions of those phases are explained in Section 4.2.

DR3. Guide children rather than judge. The professionals also highlighted that a common mistake parents make is their tendency to judge their children's behaviors and emotions. This is because parents often want to understand the context of what happened so that they can provide appropriate support or solve the issues for their children. Reflecting parental needs, some existing technology for children's emotion regulation focuses on detecting emotions by confirming with children about their emotions (*e.g.*, "*You look sad. Is that right?*" [73]). In contrast, we aimed to elicit children's emotions by acknowledging them rather than judging them. Our focus was on empathizing with emotions associated with a specific event and guiding them to develop strategies for managing their emotions by themselves first, then requesting support from their parents. Therefore, we gave specific instructions to ChaCha, focusing on acknowledging child users before intervening (*e.g.*, "`Empathize the user's emotion by restating how they felt and share your own experience that is similar to the user's.`"). Once ChaCha acknowledges children's emotions, ChaCha would guide child users through a structured conversation flow, which will be explained in detail in the following sub-section.

4.2 Conversation Design

ChaCha's conversation is designed as a state machine [83], where the system stays in one of the predefined phases with dedicated goals. Figure 3 illustrates ChaCha's conversational phases with their sub-goals and conditions for phase transition. When a new user message is added, the system analyzes the dialogue to check if the conversation met the goal of the current phase (T1–T4 in Figure 3) and forwards itself to the next phase if the goal is met. In the following, we cover each conversational phase in detail:

The Explore phase is the initial stage that the user interacts with ChaCha. This phase aims to build rapport with the user and elicit a key event and any cues for associated emotions from the user. With the user's name and age as *a priori* information (See Figure 3, left), ChaCha begins the conversation by introducing itself that it is the same age as the user, for example, "`Hi James! Nice to meet you!😊 I'm ChaCha, also 9 years old. I love to play soccer and read comics. How about you? What do you like to do?`" The introduction is prompted by an abstract instruction like "`Explain who you are and share your interests and stories`" so that the content changes every time. If the user shares a key event (See Figure 3-T1), the system transitions into the Label phase.

In **the Label phase**, ChaCha follows up the key event and helps the user label their associated emotions. The activity in this phase was inspired and informed by emotion coaching [31] and emotion card activities, where children pick emotion keyword cards

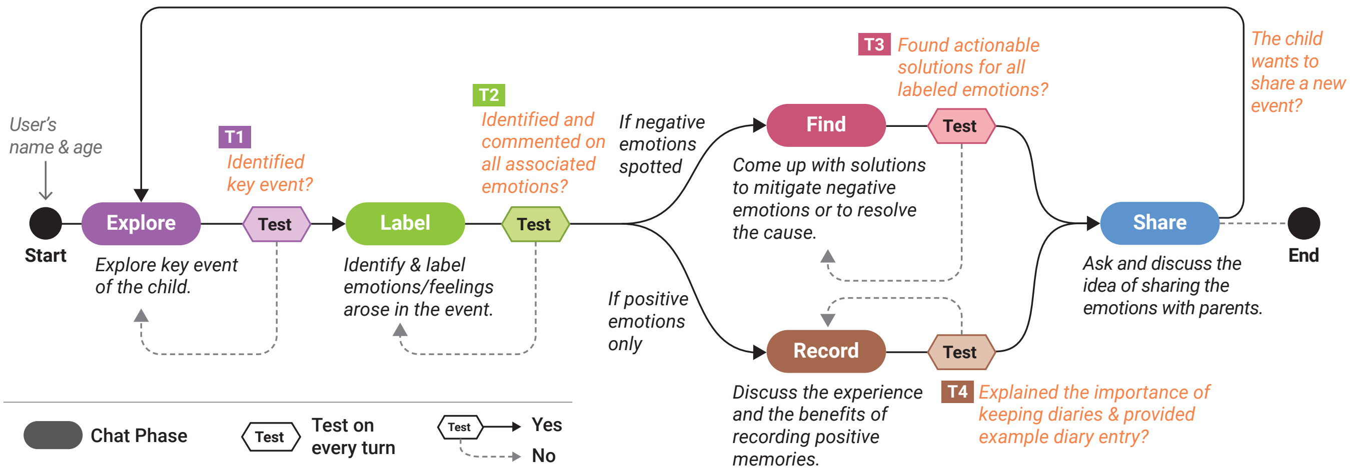

Figure 3: The overview of conversational phases of the ChaCha dialogue system and transition rules among them. Each time the user enters a message, the system inspects the entire dialogue history by performing a test corresponding to the current phase to decide whether to proceed to another phase or stay.

to describe their emotions better (*c.f.*, Figure 2b). ChaCha first probes the user to elaborate on how they felt in an open-ended way (*e.g.*, "*Can you share a bit more about how you felt at that time?*"). If the user struggles, ChaCha then provides an emotion picker (Figure 1-C) where the user can select multiple emotion labels. We curated 20 emotions from Plutchick's wheel of emotions [68] and commercial emotion cards. Those 20 emotions were presented in text labels and related emojis. The second author, a child and adolescent psychiatrist, reviewed both labels and emojis. We note that children can interpret emojis differently, depending on their age [16]. However, for ChaCha, we paired text description with emoji so that emoji are only used as supplementary information. We iterated the list with two of our formative interviewees. Once the key emotions are identified, either *in vivo* words from the user or from the emotion picker, ChaCha is prompted to acknowledge them and adequately empathize with the user (*c.f.*, DR3). If the key emotions are identified and ChaCha has acknowledged them all (See Figure 3-T2), the system transitions into the subsequent phase; if any of the emotions are marked as negative by the system, it shifts to the Find phase, and otherwise to the Record phase.

The Find phase is designed to explore actionable solutions for mitigating the negative emotional experiences identified from the prior phase or addressing the issues that provoked the emotions. Derived from emotion coaching [31], we herein define "actionable solutions" as child-developed approaches for alleviating or avoiding situations where children may experience negative emotions. We do not consider these solutions to be answers. Instead, we expect children to reflect on their experiences and emotions related to a specific event while contemplating how they would feel better if they faced the same situation. Rather than merely providing solutions, ChaCha guides children to explore potential solutions first to develop their skills to handle negative emotions. Given the importance of understanding other people's emotions in improving emotional intelligence [3, 29], if the key event involves other people, ChaCha explicitly asks the user how those people would feel.

On the other hand, **the Record phase** focuses on encouraging and introducing the benefits of recording positive memories. This approach was inspired by Carter *et al.*'s study about children's positive event diary. The diary intervention reminded children of positive experiences and boosted their well-being in the classroom by encouraging them to perceive happiness [12]. To help children understand how to write a diary about their emotions, ChaCha also provides a sample diary that summarizes their emotions and events described in the dialogue. After these phases, the system transitions into the Share phase.

In **the Share phase**, ChaCha probes the user if they have already shared their emotions and related events with their parents. If so, ChaCha compliments them and asks what happened after sharing. If not, it explains how sharing their emotion would benefit them and encourages them to share with their parents. Finally, ChaCha checks if the user has another event to tell, then shifts to the Explore phase for a new event or ends the conversation.

4.3 Dialogue System with LLMs

As long and complex prompts tend to drop the task performance of LLMs [8, 80, 85], we used dedicated prompts for each phase instead of combining instructions for all phases in a single prompt [85]. By splitting prompting by phase, we aimed to steer the LLM to follow our task instructions with a shortened input. Figure 4 above illustrates how we prompt underlying LLMs to carry on our state-based conversational flow. For each turn, one of the underlying LLMs (Figure 4-G) generates a message of ChaCha, prompted with an **instruction** (Figure 4-F) concatenated with the current dialogue history ((Figure 4-B). The instruction consists of a phase-specific part (See Appendix A) and the general speaking rules (See Appendix B) shared among phases. The general speaking rules define conversational styles such as language, tone, and length. As a safety consideration for child participants, we created a backup phase, "Help", dedicated to monitoring user's responses to detect

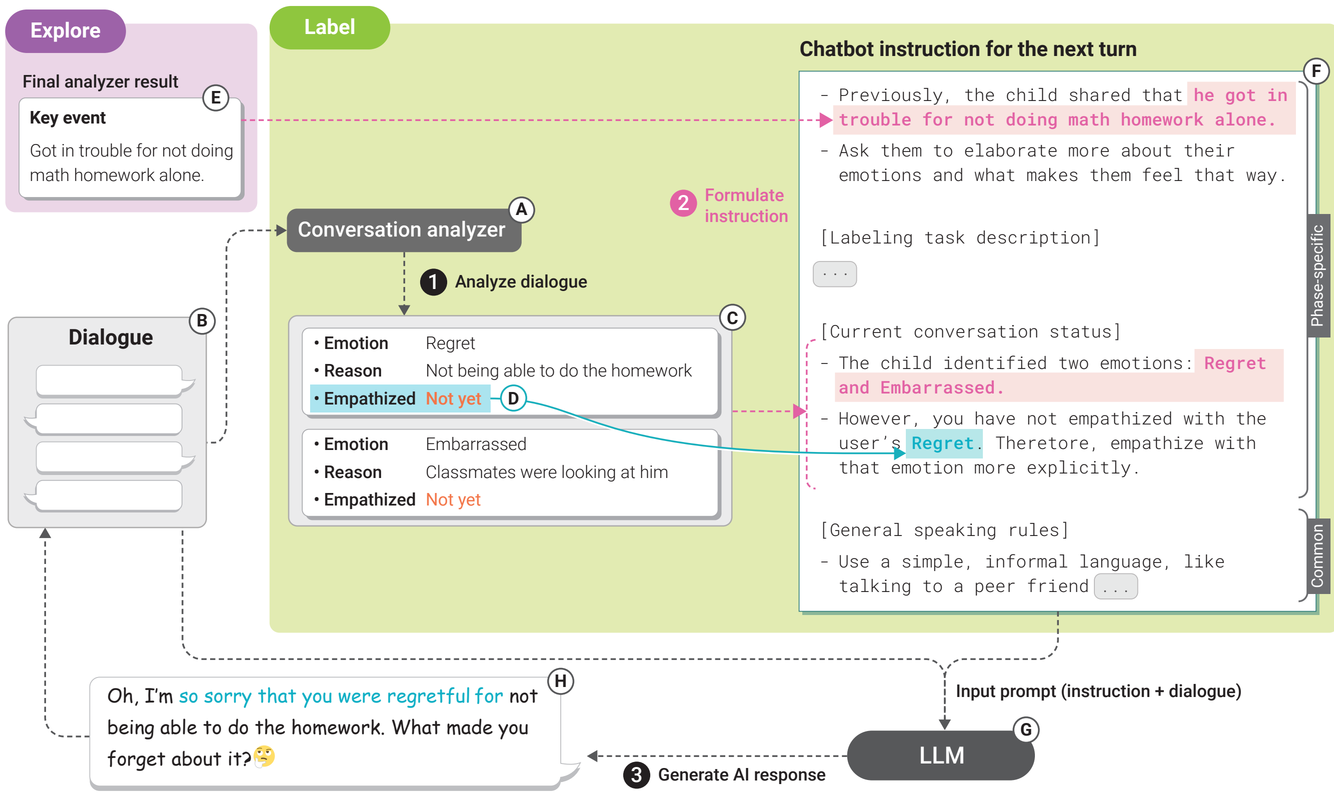

Figure 4: An example case for the mechanism of ChaCha response generation in the Label phase, especially how the LLM is prompted dynamically. Receiving the child's message, (1) the conversation analyzer (Ⓐ) analyzes the current dialogue (Ⓑ) and extracts a structured summary (Ⓒ) of what emotions are identified and whether ChaCha has acknowledged them. Combining the incomplete piece of the summary (Ⓓ) as well as the summary data from the previous phase (Ⓔ), (2) the system formulates a new instruction (Ⓕ) for the response generation. (3) That way, the LLM (Ⓖ) generates a response (Ⓗ) explicitly steered to empathize with the child's regretful event.

any indication of self-harm or suicide. During our study, none of the participants experienced the Help phase.

The phase-specific part is formulated dynamically based on the current conversation: The phase-specific **conversation analyzer** with another underlying LLM (Figure 4-Ⓐ) analyzes the current dialogue history (Figure 4-Ⓑ) and extracts a structured summary (Figure 4-Ⓒ), which contains information related to tests for phase transition (*c.f.*, T1–T4 in Figure 3). If any piece of the summary is incomplete (*i.e.*, one of the tests fails), the system includes a specific instruction that requests the LLM to generate a response in a way that complements the incomplete parts of the summary. For example, the condition to pass the test of the Label phase is to have all identified emotions acknowledged by ChaCha (See T2 in Figure 3). In Figure 4-Ⓒ, the emotion "Regret" has not been empathized yet. So the system formulates specific instruction that explicitly asks ChaCha to empathize with the regret (Figure 4-Ⓓ).

4.4 Chat User Interface

We designed ChaCha as a standard mobile text-based chat app with typing (Figure 1). A user starts a new conversation session by entering their name and age (Figure 1-A). On the input panel (Figure 1-B, bottom), the user can type in a message. The return key of the virtual keyboard enters a line break to the message, and the user should explicitly tap on the Send button (See Figure 1-B, bottom) to submit the message. This follows the typical behavior of mobile messenger apps, allowing users to enter long multiline messages. When the user enters a new message, the input panel is disabled until the system responds.

In the early stage of this work, we also considered utilizing speech interfaces, such as a smart speaker or voice chat app. Yet, we decided to use basic typing interfaces and touch interactions to minimize potential privacy concerns of children when verbalizing their private stories [22, 49]. We note that ChaCha dialogue core is UI-independent and can be easily implemented as other form factors or in other modalities.

4.5 Implementation

We implemented the core chatbot mechanism in Python, serving REST APIs for chat on a FastAPI [25] server. The chat interface, written in TypeScript [57] upon React.js [56], interacts with the server by exchanging messages between the user and the chatbot on the server. To run underlying LLMs, we leveraged OpenAI [62]'s

GPT-4 ChatCompletion API [2]. After comprehensive testing with the chatbot, we chose the `GPT-4-0613` model for response generation, considering the model's performance in Korean (*c.f.*, [61]) and token size limit [66].

We also implemented the dialogue analyzers with LLMs as they are performant at text summarization [8]. We formulated each analyzer's instruction prompt in a chain-of-thought style [81] and provided several examples to trigger few-shot learning. We used faster `GPT-3.5-turbo-0613` for dialogue analyzers to reduce the latency of the chatbot responses.

5 USER STUDY

We conducted an exploratory user study in South Korea to examine how ChaCha elicits emotional expression from children and how children interact with ChaCha. To minimize any potential threats caused by exposing children to LLMs, we conducted the study in a lab under the supervision of a researcher. Our study protocol was approved by the public institutional review board of the Ministry of Health and Welfare of South Korea.

5.1 Participants

We recruited 20 children (C1–20; nine girls) by advertising our study on a local community platform in South Korea. Following the IRB guidelines, the advertisement and screening questionnaire were targeted at the parents of potential study participants. Still, the screening questionnaire asked the respondents to confirm their children's willingness to participate. Our inclusion criteria were (1) a child aged between 8 and 12, (2) who does not have any difficulties in typing on smartphones, and (3) who can visit our lab in person. We note that the age range we set is often considered a distinctive, homogeneous stage in developmental theories (*e.g.*, psychosocial development [24], emotion regulation development [32], cognitive development [39, 67], moral development [50]). Before entering the adolescent stage in which children experience significant brain development, children ages 8–12 often experience similar cognitive and psychosocial growth [24]. To that end, this study aims to explore the broader perspectives of children in the same developmental stage rather than building personalized or customized chatbots for each age group. Table 1 summarizes the demographic information of participants. Participants' ages ranged from 8 to 12 (M = 9.75). All but five participants had their own smartphones, but the other five also used their parents' smartphones for playing games or watching videos.

5.2 Study Setup and Procedure

Each child participant went through a 1-hour study session in our lab, visiting with their parent. The study session consisted of (1) briefing, (2) conversation with ChaCha, and (3) debriefing. The parent participated in a briefing session only with their child until signing the consent forms.

Briefing. We first explained the study overview to the participants in age-appropriate language. In addition to the overall procedure of the study, we ensured that all participants understood that they would chat with a conversational AI that is configured to be the same age as them. We then attained their signed assent (and consent from the parent). After signing the assent form, the participant completed a questionnaire on the Emotional Expression Scale for Children (EESC) [64]. We used the validated Korean version of EESC [14] with 12 items on a 5-likert scale. The scores of six items indicate the child has emotional awareness of hardship, while the other six show emotional expression. The higher score indicates that children are more likely to experience difficulty in identifying and/or expressing emotions. These scores were collected to estimate each participant's emotional competencies.

Conversation with ChaCha. After completing the questionnaire, we provided the participant with a smartphone (Samsung Galaxy S22 or iPhone 13 as their preference) for interacting with ChaCha. We asked them to freely chat with ChaCha as much as they wanted. If thirty minutes passed since the participant began the chat, we asked them to end the conversation. We also ensured participants that the conversation with ChaCha would only be shared among the researchers for research purposes and never with their parents or healthcare providers unless the conversation indicates immediate mental support. For each study session, we video-recorded the participants' facial expressions to capture non-verbal expressions of their emotions while interacting with ChaCha. We also recorded the device screen to capture how participants type.

Debriefing. After chatting with ChaCha, we had a short debrief about their interaction with ChaCha. The debrief lasts between 5 to 20 minutes. The debrief questions pertained to how they perceived

Table 1: Demographic information of the participants. "Awareness" and "Expression" indicate the dimensions of the Emotion Expression Scale for Children; the averaged scores of the subscales ranged from 1 to 5. The higher score indicates the child is more likely to experience difficulty in emotional awareness and/or expression.

Alias	Age	Gender	Own smartphone	EESC	
				Awareness	Expression
C1	11	Boy	No	2.33	2.00
C2	8	Boy	No	1.83	1.83
C3	12	Boy	No	2.33	4.00
C4	11	Girl	Yes	1.67	2.33
C5	8	Boy	No	2.83	2.83
C6	8	Girl	Yes	2.67	2.00
C7	10	Girl	Yes	2.50	2.50
C8	8	Boy	No	2.67	2.33
C9	9	Girl	Yes	2.17	1.67
C10	10	Girl	Yes	1.50	1.83
C11	12	Boy	Yes	3.50	3.17
C12	11	Girl	Yes	4.00	3.67
C13	9	Boy	Yes	1.83	1.50
C14	10	Boy	Yes	1.67	2.33
C15	10	Boy	Yes	3.67	2.33
C16	8	Girl	Yes	1.50	2.00
C17	10	Boy	Yes	2.60	2.33
C18	12	Girl	Yes	1.67	1.50
C19	9	Boy	Yes	2.50	3.33
C20	9	Girl	Yes	2.00	1.33

[2]https://platform.openai.com/docs/guides/gpt/chat-completions-api

ChaCha and its conversation skills and how they felt sharing their stories and emotions with ChaCha (See supplementary material for the interview guide). We also audio-recorded each debrief for transcription and analysis. After the session, we offered participants a gift of 30,000 KRW (approx. 25 USD) as compensation.

5.3 Data Analysis

We first analyzed participants' responses to explore what events and emotions children shared with ChaCha. We reviewed dialogues to identify extracted key events and associated emotions (*c.f.*, Table 2) that participants shared. We treated only incidents that actually happened to the participant as key events. For example, we did not treat statements about a participant's hobby (*e.g.*, "*I like solving a Rubik's Cube*") as events but treated a specific occasion of hobby (*e.g.*, "*I solved a Rubik's Cube today*") as events. We report the key events and emotions in Section 6.1.

We then explored the conversational behaviors of ChaCha and participants in each phase to evaluate ChaCha's ability to steer conversation flow. We collected dialogues from the chat sessions in a spreadsheet for analysis. Hereinafter, we define *turn* as a message exchanged at a time, *user turns* as turns by the participants, and *system turns* as turns by ChaCha. We then used Pandas Python package [82] to compute diverse descriptive statistics from the collected dialogue dataset, such as the total number of turns per session. As a proxy indicator of the message length, we calculate the number of syllables by counting Korean letters only in messages, excluding non-Korean characters, symbols, and emojis. We also categorized the conversation turns with bottom-up codes. Two researchers separately coded the dialogues of two participants (10%). The coded categories were discussed, merged, and removed until the researchers reached an agreement on the final list of eight categories: *Ask/Express experience*, *Ask/Express emotion*, *Ask/Express opinion*, and *Ask/Express traits*. The dialogue data of the remaining participants were coded based on the eight categories. Due to the nature of the natural conversation, we multi-coded each turn. As a result, we coded about 90% of the dialogue turns with one or more of these eight categories. The remaining dialogue turns were short answers from participants (*e.g.*, "*Yes*", "*No*"), closing statements (*e.g.*, "*bye*", "*see you later*"), or general queries not relevant to the study context (*e.g.*, "*Do you know Siri?*"). We report the conversation patterns between children and ChaCha in Section 6.2.

We further analyzed debriefing transcripts and supplementary information from video recordings and screen captures to identify children's perspectives and expectations on ChaCha. We coded and analyzed the debriefing transcripts with Thematic Analysis [7]. Through a series of group discussions, we compared, discussed, and revised the recurring themes until the agreements were reached. We identified three salient themes: how the participants perceived ChaCha, how they felt when they shared their stories with ChaCha, and how they thought of sharing emotions with ChaCha. As for the qualitative analysis of the dialogues, we also performed open coding on the dialogues to identify recurring patterns and themes in participants' conversations with ChaCha. Those themes were then used to support the analysis of debriefing transcripts. Leveraging data from dialogue and debriefing, we found emerging patterns in participants' conversations with ChaCha and their perceptions of ChaCha as a chatbot that encourages sharing emotions. We report the children's perspectives of ChaCha in Section 6.3.

5.4 Ethical Considerations for Children

We carefully took ethical actions for child participants in our user study. First, we obtained separate consent forms from the children themselves and their parents. We further confirmed each child participant's desire to participate in the study before each study session. Second, we reminded the children that they could stop participating in the study whenever they felt uncomfortable. We informed them that, during each session, a researcher would be in the study room with the participant to halt the study upon request or provide immediate support. Third, before each session, we confirmed with the parents that their children did not have any mental health issues. We also ensured all children acknowledged that they would interact with a chatbot, not a real person.

6 RESULTS

From the user study, we collected 20 dialogues comprising 878 turns (434 user and 444 system turns), with 43.9 turns per session on average (SD = 15.22, min = 27 [C17], max = 87 [C18]). On each turn, participants included 18.08 syllables per message on average (SD = 8.93, min = 3.25 [C14], max = 33.68 [C10]), while ChaCha had 82.06 syllables on average (SD = 21.89, min = 48.57 [C5], max = 130.56 [C1]). On average, participants took 66.85 seconds to type and send one message (SD = 27.59, min = 20.5 [C14], max = 113.67 [C9]).

In this section, we report the results of our user study in three parts to demonstrate the feasibility of an LLM-driven chatbot in prompting children to share their emotions about personal events. In Section 6.1, we provide the overview of key events and emotions that child participants shared with ChaCha. In Section 6.2, we explore how ChaCha and the children conversed about key events and associated emotions, focusing on the patterns of exchanging questions and information in different conversational phases. Lastly, in Section 6.3, we report on the children's perceptions of ChaCha and their overall experience of sharing emotions with ChaCha based on the debriefing.

6.1 Shared Key Events and Emotions

Participants shared various types of key events and related emotions (See Table 2). Those events generally fell into four types: Recent trips (*e.g.*, theme park, family trip), Personal achievements (*e.g.*, 1st place in a race, riding a subway train by oneself), Concerns (*e.g.*, procrastination on homework, conflicts with mom), and Hobbies (*e.g.*, maze puzzle, Rubik's cube). Those events were identified as "key" events since participants shared specific emotions related to each of them.

Some participants shared events that induced negative emotions (*e.g.*, fear). In such cases, ChaCha prompted participants to share their opinions and develop ideas for potential solutions to manage such emotions (See Table 2). As described in Section 4.2, "solutions" indicate the ways to alleviate or avoid the events where children had already experienced negative emotions. All solutions, except C3 and C11, were initially developed by the participants after reflecting on their emotions and the key events prompted by ChaCha. Among

Table 2: Key events that the child participants shared, emotions identified from conversations, and solutions to resolve negative emotions. The 💬 symbols indicate the emotions directly described by the participant, whereas the ☑ symbols indicate those selected from an emotion picker (See Figure 1-C).

	Key Event	Emotions identified	Solutions to negative emotions
C1	Solved a maze puzzle	💬 Accomplished 💬 Simple pleasures	
	Snorkeled when visited Palau for a family trip	💬 Epic 💬 Want to visit again 💬 Beautiful	
C2	In a game, he killed another character and collapsed.	💬 Fun ☑ Surprise	In a game, carefully look around for other people.
C3	Rode a roller coaster in a theme park	💬 Fear ☑ Satisfaction ☑ Surprise	Since riding a roller coaster should be scary, so C3 wanted to ride it as it is
	Rode a water slide	💬 Thrilled 💬 Fun 💬 Refresh 💬 Fear	C3 liked it as it is
	Rode a gyro drop for the first time in Vietnam	💬 Fear ☑ Joy ☑ Passion ☑ Surprise	Try higher gyro drop next time
C4	Came across a retired math teacher	☑ Joy ☑ Surprise	
C5	Watching Youtube		
	Rode the Pirate Ship in a theme park	💬 Fun	
C6	Made a pizza with her mom	💬 Happy	
C7	Won the 1st place in a running race	💬 Joy ☑ Accomplished	
	Read a book about laws	💬 Surprised	
C8	Could not solve Rubik's cube	☑ Confidence ☑ Annoyance	Study more about how to solve a Rubik's cube
C9	Cooked and had instant noodles with her friends when her mom was absent	☑ Joy ☑ Thrill ☑ Happy ☑ Satisfaction ☑ Accomplished	
	Rode subway by herself	💬 Accomplished ☑ Confidence	
C10	Performed at an orchestra concert at school but had to move to another school	💬 Proud 💬 Shame	Contacting friends from the old school and making new friends at the new school
	Went to a water park	💬 Pleasure 💬 Happy	
C11	Had difficulty in solving math problems and got in trouble with his mom	💬 Sad 💬 Challenging ☑ Fear ☑ Distress	Solve math problems on a paper, and review the problems
C12	Procrastinated on doing summer homework	💬 Apprehensive	Cramming homework
	Shared the story about Oppenheimer with CC	💬 Tragic 💬 Cool ☑ Thrill ☑ Passion ☑ Satisfaction ☑ Accomplished ☑ Regret ☑ Annoyance ☑ Disappointment	
C13	Got in trouble with his mom because he did not did well on solving math problems	☑ Regret ☑ Sorry	Solving math problems by himself for an hour every day
	Could not go swimming due to unfinished homework	☑ Disappointment ☑ Discomfort	Completing homework beforehand so that he can do what he likes (*e.g.*, swimming)
C14	When playing baseball with his friends, hit in a row	💬 Fun ☑ Joy	
C15	Found a sick bird on the street, but it was actually pooping	💬 Embarrassed ☑ Surprise ☑ Fear ☑ Regret ☑ Discomfort	
	Rode a rollercoaster in a theme park	💬 Scared ☑ Passion ☑ Confidence ☑ Fear	
C16	Cooked a pizza with her friend	💬 Fun	
	Did not have masks on the first day of school, but no one wore masks	💬 Relief	
	She likes rope jumping, but is the shortest in the class	☑ Passion ☑ Surprise ☑ Regret ☑ Resentment ☑ Envy ☑ Disappointment ☑ Discomfort	

<Continued on the next page>

<Continued from the previous page>

	Key Event	Emotions identified	Solutions to negative emotions
C17	Played soccer with friends and won a game	Happy, Joy, Thril, Happy, Passion, Satisfaction, Confidence, Comfort	
C18	Made burgers with her parents at home	Joy, Moved, Thrill, Happy, Satisfaction, Comfort	
	Went to a K-pop concert	Joy, Regret	
C19	Played badminton and had ice cream with his parents	Pleased	
C20	Likes dodgeball and Pikachu	Joy, Happy, Satisfaction	
	Went on a trip to Jeju island	Enjoyable, Tired	

those two exceptional participants, C3 refused to come up with solutions since he did not think it was necessary and believed that riding a roller coaster was supposed to be scary, so he would be scared anyway. C3's rejection indicates that he perceived being scary is not a negative emotion he needs to avoid while riding a roller coaster (the key event). In C11's case, he could not think of any solutions. Thus, ChaCha suggested several options (*e.g.*, watching videos, solving practice problems) until C11 agreed on solving math problems on a paper sheet and reviewing them.

In addition to key events, participants also shared and labeled multiple emotions for a given event. All but three (C4, C8, C13) participants could describe at least one emotion by themselves without the help of the emotion picker. In particular, some of them explained the emotions described in their own words to ChaCha. For instance, C1 mentioned feeling "simple pleasures" when solving a maze puzzle. When ChaCha asked what it means, C1 explained that simple pleasures are "a small amount of joy" that significantly impacts him. Similarly, C3 described that riding a roller coaster in a theme park was fearful. Still, instead of just saying "fear," he mentioned, "shiiiivering." which was a intended typo for emphasis on the intensity of fear. ChaCha understood what he meant and generated responses about a similar experience (*e.g.*, feeling scared when riding a roller coaster) to empathize with C3.

6.2 Exchanges of Questions and Expressions

Figure 5 visualizes the distribution of questions and expressions in four themes (experience, emotion, opinion, and trait) within each conversational phase. The x-axes indicate the ratio of the number of turns coded with the category to the total number of turns in the phase.

The distribution patterns varied depending on the goal of each phase. In the Explore phase, the even distribution of ChaCha's and participants' expressions of personal traits shows how ChaCha attempted to build common grounds with participants by stating their shared interests and hobbies. Then, in the Label phase, ChaCha focused on asking about participants' emotions. In Find, it prompts participants to consider actionable solutions (Opinion). About 30% of this phase's conversation turns were generated by participants who described potential solutions to alleviate their negative emotional experiences. Moreover, ChaCha expressed and asked a similar amount of opinions in Record since it described the benefits of capturing positive experiences as written documents and asked about participants' thoughts. Likewise, ChaCha inquired about participants' thoughts and their experiences of sharing their emotions with parents in Share.

Although this distribution of conversation turns showed how each phase successfully achieved its primary goal and guided the

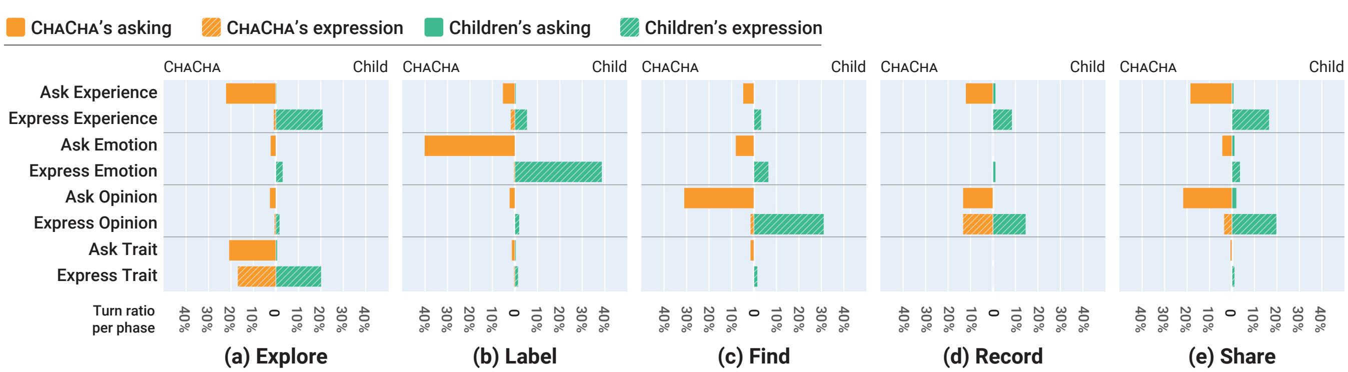

Figure 5: The frequency of the eight asking-expression categories by conversational phase. The orange bars indicate the average turn ratio for ChaCha's turns, whereas the green bars indicate those for participants' turns. Bars with a hatching pattern indicate turns coded with the expression codes.

conversation flow, ChaCha predominantly expressed traits and experiences to build rapport with the participants. Throughout the conversations, ChaCha rarely shared emotions or experiences. As we prompted ChaCha's underlying LLM mainly with high-level behavioral instructions rather than predefined contents, ChaCha shared information based on the participants' previous responses. As described in the dialogue below, ChaCha generated and shared a similar experience of riding a roller coaster for the first time and feeling scared.

C3	*(Riding a rollercoaster was) just scary. I missed my mom, that's it.*
	Yeah, it can happen in scary situations. I remember being scared and reaching out to my dad when riding the rollercoaster for the first time.

Some participants began asking ChaCha questions in the later phases (Record, Share). Those questions relate to ChaCha's preferences, experiences, concerns, emotions, and general queries about ChaCha, such as "*Why is your name ChaCha?*" According to the debriefing, participants asked questions to ChaCha (1) to build more common ground with ChaCha and (2) to listen to ChaCha's stories, just like talking with a friend. Since we did not specify how to handle such questions, ChaCha mostly said it does not know or can not share anything about its experiences.

For instance, C20 stated that she likes math and then asked ChaCha when its happiest moment was. ChaCha answered that it could not feel any emotions but was still glad to know more about C20's emotions.

C20	*Among the school subjects, I like Math the most.* ***By the way, when is your happiest moment?***
	It is incredible that you like Math the most! Math can be enjoyable and challenging. Yes, I like what you asked, but ***I can't feel the emotion because I am a computer app.*** *But I can always say I'm happy when I talk to you and learn more about your joy and happiness. Come to think of it, what kind of feelings do you feel while doing math? Can you tell me why you like math?*

As shown in the examples above, ChaCha mainly shared its experiences based on the participant's experience or acknowledged the limitation of its self-expression. However, some participants were disappointed about ChaCha's behaviors of restricting self-expression: C18 remarked in the debriefing, *"[ChaCha] would only empathize with me, so I asked it to tell me its story at the end, but it didn't, so I was a bit sad."*

On the other hand, ChaCha occasionally shared its background information that was not specified in the model instructions. C7 shared her dream job as a lawyer and then asked ChaCha about its dream. ChaCha replied that it wants to become someone who listens to and sympathizes with many people while communicating with them.

C7	*My dream is to become a lawyer*
	Wow, what a dream! A lawyer does an essential job. I don't have many friends who say they want to be a lawyer, so it sounds special. Why do you want to be a lawyer?
C7	*Because I saw on TV that some people are being treated badly, and I want to stop it.* ***Do you have a dream? Please tell me if you have!***
	I see. Your desire to be a lawyer fighting for justice is really cool. I actually have a dream, too. Since I like to communicate with many people and listen to their stories, ***I want to be a person who listens to their stories and sympathizes with them while communicating with people.*** *So, I continue to work on self-improvement. Now, shall we talk about a special moment or experience related to your dream?*

In the dialogue, ChaCha shared its dream to be "a person" who listens to people's stories. This description of ChaCha's dream was not specified in the instructions. After sharing about its dream, ChaCha smoothly steered the conversation flow back to the Label phase in which ChaCha identifies a key event and related emotions from participants.

6.3 Children's Perception of ChaCha

Based on the analysis of the participant debriefs, we identified how they perceived ChaCha's persona and conversation skills about sharing their emotions.

6.3.1 ChaCha's Persona. Most participants perceived ChaCha as a close friend with whom children can share their emotions and secrets. Some participants perceived ChaCha as older or even younger than themselves, even though we set ChaCha the same age as each participant. For instance, C1 thought ChaCha would be an older friend since ChaCha carefully listened to his stories and provided advice: *"[ChaCha] gave me some advice, and listened to my experience well, so it feels like [ChaCha] is someone a little older than me".* On the other hand, C3 felt ChaCha was younger than himself because sharing stories and emotions is what he used to do when he was younger: *"Sharing personal conversations [about emotions] is what I usually did when I was younger. [ChaCha] seems to be younger than me since it wanted to have such conversations."* Regardless of different perceptions of the age of ChaCha, almost all of the participants recognized ChaCha as a close friend with whom they could easily converse. Exceptionally, C4 experienced moments when she felt ChaCha was not an actual friend since ChaCha did not share its stories; instead, it kept asking questions. Although she felt comfortable interacting with ChaCha, she perceived that the one-way inquiry was not typical interaction with her friends.

6.3.2 ChaCha's Conversational Skills. Most participants perceived that ChaCha was good at empathizing with their emotions. Its empathizing conversation skills built participants' trust in ChaCha and encouraged them to share their stories. Participants mentioned they would share their secrets or sad feelings with ChaCha since it listens to their emotions and helps them find solutions, even though they would not share those with their parents. On the contrary, C14 mentioned that he would share his sad feelings with his mom instead of sharing with ChaCha since she would better comfort him than ChaCha: *"I don't think I would tell ChaCha [about negative* *emotions] (. . .) I think Mom would do something more helpful for me. Mom comforts me, but I don't think ChaCha can do that."* This quote indicates C14's clear boundary of sharing emotions with ChaCha, even though he had an overall positive experience from interacting with ChaCha.

Moreover, ChaCha's probing questions provoked participants to reflect on their experiences and emotions. For instance, C9 shared how ChaCha's questions related to emotions helped her remember memories associated with specific events and emotions. During the study session, she spent a longer time on typing and revising her responses to ChaCha's questions, indicating her reflection on specific events and emotions. ChaCha's conversational skills in talking about emotions also encouraged many participants to learn and apply those skills to their conversations with their family members or friends: *"I would talk more about emotions with my family members (. . .) I feel more confident after talking with ChaCha" (C9).* They mentioned that they could adopt ChaCha's conversation style when expressing their emotions to others.

6.3.3 Children's Expectations for Future Interactions with ChaCha. In the debrief, participants shared their willingness to chat more and expectations for future interactions with ChaCha. All participants, except C5, wanted to chat with ChaCha again in the future. C5 prefers to talk to his friends since ChaCha can not actually "speak" with its voice: *"I prefer my friends to ChaCha (. . .) They seem to know what I want (. . .) ChaCha can not speak, so I want to talk with my friends."* Although the modality of the chatbot was out of scope for this study, C5's quote showed that text-based interactions were inadequate to perceive ChaCha as a friend to share his emotions.

For the remaining 19 participants, the preferred frequency for conversing with ChaCha ranged from once in a day to once in a year. Some of them prefer to chat with ChaCha less frequently because it would be boring if they share the same daily routines every day: *"If I chat with ChaCha too often, I think it'll all be similar every day (. . .) I'm going back to school soon [after break], so I guess it will be similar [conversations] every day because the classes are repeated, but I think it'll be different if I take some time." (C17).* In addition, some participants prefer less frequent interactions due to the burdens of answering questions. During the study sessions, they felt burdened by the one-way process of sharing their emotions and experiences. For instance, C9 and C18 mentioned that asking about each emotion individually was annoying since they had to repeatedly provide answers about different emotions that they shared with ChaCha. It was mainly because both described and selected more emotions than other participants.

7 DISCUSSION

In this section, we discuss lessons learned from the design and implementation of ChaCha and the exploratory study, focusing on the benefits and concerns of leveraging LLMs to encourage children to share stories and emotions. We also discuss the design considerations for developing an LLM-driven chatbot for children, particularly for emotion conversations. Lastly, we report on this study's limitations and the potential directions for future work.

7.1 Benefits of Leveraging LLM to Prompt Children to Share Personal Events and Emotions

The LLM-driven chatbots are well-known for their ability to have natural conversations with users. In addition to this aspect of LLMs, we found two distinct benefits of leveraging LLMs for prompting children to share personal events and emotions.

First, LLM-driven chatbots have the potential to facilitate more empathetic conversations with children by establishing common ground. Prior studies have presented opportunities for chatbots as promising interventions in the healthcare domain; some of them are specifically developed to recognize users' emotions and deliver empathetic responses (*e.g.*, [27, 52, 54]). However, most of them are designed for adult users, predominantly focusing on detecting users' emotions through sensor data and algorithms while showing empathy. ChaCha, on the other hand, aims to encourage children to describe personal events and contemplate associated emotions. Instead of detecting emotions, LLMs enabled ChaCha to recognize the context of children's experiences better and show empathy that helps them define their emotions. As shown in the dialogues in Section 6.2, ChaCha built common ground by generating responses that share similar preferences and traits with each child (*e.g.*. felt scared when riding a roller coaster). This empathy-based common ground made children perceive ChaCha as a close friend with whom they can share their stories and emotions. The perception of ChaCha as a close friend figure to develop emotional competence resonates with the Zone of Proximal Development in education; it refers to a space between a child's ability to learn with and without guidance from a more skilled person [79]. In our study, ChaCha—a more experienced person—guided child participants—a less skilled individual—to share personal events, reflect on their experiences, and express their emotions. This relationship was evident in C9's case, where she gained confidence in sharing her emotions with others (*e.g.*, parents) and learned how to express her emotions from the conversations with ChaCha. To enhance this relationship in the Zone of Proximal Development, LLM-driven chatbots could be designed to recognize each child's emotional capabilities so that the chatbots can provide proper guidance on specific skills that the child needs to improve (*e.g.*, labeling emotions). More personalized guidance and empathetic responses would allow children to develop specific skills to label their emotions, explore potential solutions to negative emotions, and share their emotions with parents by prompting children to reflect on events they experienced. Future work is needed to explore and examine how such an empathy-based relationship between an LLM-driven chatbot and children may promote the development of children's emotional competencies in long-term intervention.

Second, LLM-driven chatbots have the potential to effectively steer the conversation flow with children's responses on various topics. In our study, participants shared various personal events and different associated emotions. Recognizing those events and emotions, ChaCha utilized LLMs as a state machine that steers the conversation flow to pursue the sub-goals at hand in each phase. As LLMs perform in-context learning, where the model "somehow" reads the latent concept in input prompts written in plain human language, it is inherently challenging for designers to fully control the conversational flow [45] or modify patterned behaviors ingrained in the model. For example, generalized chatbots like ChatGPT [60] tend to provide a long and detailed response to the first user message [53]. Such behaviors are not adequate for children [53]. Our state-based approach successfully steered an LLM to carry on a child-friendly conversation while complying with the sub-goals we defined. This ability for conversation steering is particularly essential in the emotion coaching process [31] in which distinctive steps guide children to increase awareness of their emotions and potentially others' emotions. This approach may help overcome the limitations of rule-based chatbots that require children to submit structured responses (*e.g.*, only yes or no to confirm their feelings [73]). Therefore, our study suggests that the state-based approach is a promising design method for LLM-driven chatbots, especially when designers want to apply predefined conversational goals.

7.2 Designing More Engaging, but Guiding, Chatbot Persona for Children

In our study, almost all participants highly anthropomorphized [19] ChaCha, perceiving it as a close friend. This seemed mainly due to its friend-like elements, including having common interests or empathizing with their emotions. Surprisingly, some of them even wanted to know more about ChaCha's dreams or emotions, which was beyond our expectations. In most cases, ChaCha "improvised" by providing the information missing in our instruction. As shown in the example of C7's dialogue about her dream job, ChaCha's improvised response about its dream job may provoke children's interest in getting to know more about ChaCha.

Although our participants showed high engagement and rapport with ChaCha, we also raised concerns about children's overreliance on the chatbots. With more advanced AIs, overreliance on AI has become a critical issue [63], particularly in the child-AI interaction context. Our findings showed participants' desire to share their stories and secrets that they never shared with their parents. This overreliance on ChaCha is what mental health professionals in our formative study also warned about. The professionals articulated the importance of sharing children's emotions and stories with their parents rather than dumping everything into a chatbot. They expected chatbots as a temporary tool to guide children to share their emotional needs with parents or professionals rather than replacing their roles as caregivers. Although having emotional conversations with an LLM-driven chatbot differs from completing a task with AI, similar approaches to reducing overreliance can be applied to designing a chatbot for children. One approach can be cognitive forcing interventions that compel users to engage thoughtfully with an AI system [11]. In the context of the emotional conversation with children, an AI-driven chatbot may integrate emotion-related activities (*e.g.*, role play and guessing the emotions of each character) so that children have the opportunity to reflect on their and others' emotions rather than perceive the chatbot as someone who can easily offer solutions to their emotional problems.

In sum, the future version of ChaCha or any other LLM-driven chatbots for children could incorporate a more concrete yet guiding profile of the chatbot's persona. Then, this guiding persona can be consistently referred to throughout the interaction with children. Natural language processing techniques such as persona attribute extraction (*e.g.*, [84, 89]) can help extract the child user's persona attributes from the ongoing dialogue and use them to configure the chatbot's persona. In that way, ChaCha could adapt to children with different personality traits and interests.

7.3 Considerations for Children's Long-term Interactions with an LLM-driven Chatbot

Our study showed the opportunity for an LLM-driven chatbot to facilitate emotional conversations with children. Importantly, 19 out of 20 study participants expressed a willingness to have a longer engagement with ChaCha in the future, while their preferences for engagement frequency varied. As shown in the subsections above, the long-term use of ChaCha may amplify the benefits of leveraging LLMs (*e.g.*, building rapport with children). However, based on our findings, we also identified potential issues regarding inconsistent and harmful messages that may occur when children interact with an LLM-driven chatbot across multiple sessions.

First, more prolonged interactions may result in a potential breakdown of behaviors or character profiles of an LLM-driven chatbot. Our findings showed that ChaCha sometimes improvised responses to participants' questions (*e.g.*, "*What is your dream?*"). Since this improvisation was not in our instructions, its answer to the question may be inconsistent if the same question is asked again later (*e.g.*, ChaCha's dream keeps changing). A consequence of such inconsistency can be the decrease in children's engagement with ChaCha since children would no longer consider ChaCha as someone to trust or share their emotions. To prevent this issue, a more thorough construction of the chatbot's character profile will help the chatbot maintain consistency and retain the children's engagement. For more engaging and effective conversation, the profile can be personalized, reflecting various aspects of the child user, such as their emotional development status or relationship with parents.

Second, long-term use of LLM-driven chatbots may increase the chance of generating responses that can potentially be harmful to children. There can be two cases where a chatbot's responses can be detrimental: (1) When the message itself is toxic (*e.g.*, violent language) or (2) a response can be implicitly harmful in a specific context or a child's unique conditions (*e.g.*, a child with a traumatic event). The former can be addressed by fine-tuning or incorporating additional response filters trained with harmful and toxic language datasets (*e.g.*, [43, 48]). The latter demands a more sophisticated approach, such as constructing a multi-session chat dataset between a child and an AI with improper messages annotated by child mental health professionals.

7.4 Incorporating Parents in Child-Chatbot Interactions

Drawing on child mental health professionals' comments and the user study findings, we learned that incorporating parents' needs is necessary for child-chatbot interactions. While we do not urge restrictive parental control over children's chatbot use, we suggest chatbots for children should consider how parents expect chatbots to behave and how they may intervene in child-chatbot interactions. Especially, addressing potential privacy tensions between children and parents will be critical. The tension between online safety and parental surveillance in mobile apps for children has already been discussed in the CHI community (*e.g.*, [28]). Such tension may also occur in child-chatbot interactions. Our findings showed that some participants prefer to share their stories and emotions only with ChaCha rather than their parents. This preference for a private conversation with ChaCha may cause privacy tensions between children and parents due to their potentially conflicting perspectives. In the formative study, some professionals shared their experiences with specific cases where parents keep asking their children questions to understand their children's situations or problems. On the contrary, children prefer acknowledgment of their feelings rather than an investigation into what happened. This conflict may be amplified if children privately share their emotions with a chatbot.

Hence, we believe the development of LLM-driven chatbots for children should incorporate parents' expectations and potential concerns, even if they are not primary end-users. Children aged 8–12 still need parental guidance in developing their emotional competencies. Thus, we envision chatbots to draw on parents' input about their expectations for what skills chatbots should encourage children to learn more. Ultimately, chatbots should facilitate more engaging emotion communication between children and parents rather than cause tensions between them. Like how we designed ChaCha's persona and conversation flow to guide children, LLM-driven chatbots should clarify their supplementary roles to support children rather than replace any support from parents or healthcare professionals. Further study is needed to explore the opportunities of LLM-driven chatbots to incorporate parents' voices on how their children should interact with the chatbots.

7.5 Technical Considerations for LLMs and Chatbot Language

In this section, we note some technical considerations for using LLMs and the choice of chatbot language to design a chatbot for children. Despite the distinct benefits of LLMs, we found potential drawbacks of leveraging LLMs to facilitate children's emotional conversations. Given the uncertain nature of how LLMs work, we encountered several instances where the LLM did not follow the guidelines of the current phase (*e.g.*, displaying an emotion picker out of the Label phase), or in the second or third cycles of the phases (*i.e.*, after going back to the Explore phase to share new events) where the dialogue history became long. We suspect that a long dialogue history impacted the response generation more than the provided instruction did, not achieving the sub-goals in each phase. Although the user may overlook such flaws as the conversation flows organically, the future design of ChaCha may involve additional classifiers to double-check whether the generated response is adequate for the current phase.

Moreover, we needed to design ChaCha to speak Korean. Thus, the considerations for selecting an LLM model included the model's accessibility, multilingual dialogue performance, and tokenization (translation of human language into a machine language) efficiency. We decided to use GPT-4, the most capable LLM publicly accessible via commercial API and with viable Korean dialogue generation performance (*c.f.*, [61]). Yet, tokenization of GPT is known to be imbalanced, producing about 3 to 5 times more tokens from the Korean text than the English with the same content [46, 66]. This means that Korean dialogues reach the inherent token size limit of GPT way faster than English dialogues. Therefore, the language constrained our model choice to select the most capable model with the longest token size limit (*c.f.*, GPT-3.5's token limit is 4096, and GPT-4's token limit is 8192). Thus, we believe that a cheaper and faster model like GPT-3.5 would still yield an equivalent quality of ChaCha's conversation performance if applied to languages with fewer tokens, especially English.

7.6 Limitations and Future Work

Our study has some limitations to note. First, our study population is limited to Korean children. Cultural bias may exist in how children express their emotions. Also, smartphone ownership of children is relatively high in Korea. According to a study, about 72% of Korean children owned a smartphone by ages 11 to 12 [42]. Thus, Korean children are more likely to be familiar with smartphone chat apps. Children in different cultural contexts may have other interactions and perceptions of an LLM-driven chatbot. Second, our chatbot used the Korean language on the LLM trained with a less Korean dataset. Although all outputs of the chatbot were understandable to children, some phrases were awkward and took longer to generate since the provided instructions in Korean take more tokens than those in English when delivering the same content. Third, due to the small sample size, our findings could not present a meaningful pattern in gender, which is a key factor in emotion communication (*e.g.*, [5, 51]). Future works should explore potential gender differences in children's perception of LLM-driven chatbots. Lastly, since our study was conducted in lab-based settings, children may experience different interactions with the chatbot when they use it in the wild (*e.g.*, longer-term conversations that allow the chatbot to follow up with children's previous emotions). Therefore, future work can investigate how children interact with LLM-driven chatbots to share their stories and emotions in more natural settings in the long term.

8 CONCLUSION

In this study, we designed an LLM-driven chatbot and conducted a user study with 20 children (ages 8–12), examining how they shared personal events and described related emotions with the chatbot. We reported three contributions: (1) We designed a novel LLM-driven chatbot, ChaCha, that helps children share their stories and emotions through free-form conversations; (2) We provide an understanding of how children interact with an LLM-driven chatbot in the context of story and emotion sharing; and (3) We present the benefits of leveraging LLM in designing chatbots for children and suggestions to mitigate potential risks. Drawing from our findings, we invite researchers in the CHI community to further examine the potential use and concerns of leveraging LLMs to design child-friendly chatbots.

ACKNOWLEDGMENTS

We thank our child participants and their parents for their time and efforts. The conversational flow design of ChaCha is grounded in Woosuk Seo's dissertation research, which was supported by National Science Foundation CAREER Grant #1942547 (PI: Sun Young

Park). We are particularly grateful for the insights provided by Sun Young Park and Mark S. Ackerman, which have been instrumental in shaping this work. We also thank Eunkyung Jo and Kyungah Lee for their feedback on the early version of this paper and the chatbot prototype, respectively. This work was supported by NAVER AI Lab through a research internship.

REFERENCES

[1] Alaa A Abd-Alrazaq, Mohannad Alajlani, Ali Abdallah Alalwan, Bridgette M Bewick, Peter Gardner, and Mowafa Househ. 2019. An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics* 132 (2019), 103978.

[2] R Abdelghani, YH Wang, X Yuan, T Wang, H Sauzéon, and PY Oudeyer. 2022. GPT-3-driven pedagogical agents for training children's curious question-asking skills. ArXiv. *preprint arXiv* 2211 (2022).

[3] Alberto Alegre. 2011. Parenting Styles and Children's Emotional Intelligence: What do We Know? *The Family Journal* 19, 1 (2011), 56–62. https://doi.org/10.1177/1066480710387486

[4] Godwin S Ashiabi. 2000. Promoting the emotional development of preschoolers. *Early childhood education journal* 28 (2000), 79–84.

[5] Ana Aznar and Harriet R Tenenbaum. 2020. Gender comparisons in mother-child emotion talk: A meta-analysis. *Sex roles* 82 (2020), 155–162.

[6] Byung-yeul Baek. 2023. Naver's HyperCLOVA X offers enhanced AI experience. https://www.koreatimes.co.kr/www/tech/2023/09/129_357754.html

[7] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems (NeurIPS '20)*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[9] Amy Bruckman, Alisa Bandlow, and Andrea Forte. 2007. HCI for Kids. In *The human-computer interaction handbook*. CRC Press, 819–836.

[10] Laura E Brumariu and Kathryn A Kerns. 2015. Mother–child emotion communication and childhood anxiety symptoms. *Cognition and Emotion* 29, 3 (2015), 416–431.

[11] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21.

[12] Paul J Carter, Brendan Hore, Leona McGarrigle, Manon Edwards, Gavin Doeg, Rachel Oakes, Aisling Campion, Grace Carey, Katie Vickers, and John A Parkinson. 2018. Happy thoughts: Enhancing well-being in the classroom with a positive events diary. *The Journal of Positive Psychology* 13, 2 (2018), 110–121.

[13] Siyuan Chen, Mengyue Wu, Kenny Q Zhu, Kunyao Lan, Zhiling Zhang, and Lyuchun Cui. 2023. LLM-empowered Chatbots for Psychiatrist and Patient Simulation: Application and Evaluation. *arXiv preprint arXiv:2305.13614* (2023).

[14] Eunsil Choi and Bang HeeJeong. 2011. Reliability and Validity of the Emotion Expression Scale for Children. *The Korean Journal of Developmental Psychology* 24, 3 (2011), 105–128.

[15] Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. 2022. PaLM: Scaling Language Modeling with Pathways. https://doi.org/10.48550/ARXIV.2204.02311

[16] N da Quinta, E Santa Cruz, Y Rios, B Alfaro, and I Martinez de Marañón. 2023. What is behind a facial emoji? The effects of context, age, and gender on children's understanding of emoji. *Food Quality and Preference* 105 (2023), 104761.

[17] Elizabeth L Davis, Linda J Levine, Heather C Lench, and Jodi A Quas. 2010. Metacognitive emotion regulation: children's awareness that changing thoughts and goals can alleviate negative emotions. *Emotion* 10, 4 (2010), 498.

[18] Susanne A Denham and Sharon Auerbach. 1995. Mother-child dialogue about emotions and preschoolers' emotional competence. *Genetic, social, and general psychology monographs* 121, 3 (1995), 311–337.

[19] Ameet Deshpande, Tanmay Rajpurohit, Karthik Narasimhan, and Ashwin Kalyan. 2023. Anthropomorphization of AI: Opportunities and Risks. arXiv:2305.14784 [cs.AI]

[20] Gilly Dosovitsky and Eduardo Bunge. 2023. Development of a chatbot for depression: adolescent perceptions and recommendations. *Child and Adolescent Mental Health* 28, 1 (2023), 124–127.

[21] Michael Dowling and Brian Lucey. 2023. ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters* 53 (2023), 103662.

[22] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human-Computer Interaction* 31, 4 (April 2015), 307–335. https://doi.org/10.1080/10447318.2014.986642

[23] Suzanne M Egan, Jennifer Pope, Mary Moloney, Clara Hoyne, and Chloé Beatty. 2021. Missing early education and care during the pandemic: The socio-emotional impact of the COVID-19 crisis on young children. *Early Childhood Education Journal* 49, 5 (2021), 925–934.

[24] Erik Erikson. 1959. Theory of identity development. *E. Erikson, Identity and the life cycle. Nueva York: International Universities Press. Obtenido de http://childdevpsychology. yolasite. com/resources/theory% 20of% 20ident ity% 20erikson. pdf* (1959).

[25] FastAPI. 2023. FastAPI framework, high performance, easy to learn, fast to code, ready for production. Retrieved Aug 25, 2023 from https://fastapi.tiangolo.com/

[26] Amy L Gentzler, Josefina M Contreras-Grau, Kathryn A Kerns, and Barbara L Weimer. 2005. Parent–child emotional communication and children's coping in middle childhood. *Social Development* 14, 4 (2005), 591–612.

[27] Asma Ghandeharioun, Daniel McDuff, Mary Czerwinski, and Kael Rowan. 2019. Emma: An emotion-aware wellbeing chatbot. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 1–7.

[28] Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J. LaViola Jr, and Pamela J. Wisniewski. 2018. Safety vs. Surveillance: What Children Have to Say about Mobile Apps for Parental Control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Montreal QC</city>, <country>Canada</country>, </conf-loc>) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173698

[29] Daniel Goleman. 2005. *Emotional Intelligence: Why it Can Matter More Than IQ* (10th anniversary ed.). Bloomsbury. 384 pages.

[30] Google, Inc. 2023. Bard - Chat Based AI Tool from Google, Powered by PaLM 2. Retrieved Aug 25, 2023 from https://bard.google.com/

[31] J.M. Gottman and J. Declaire. 1998. *Raising An Emotionally Intelligent Child*. Simon & Schuster. https://books.google.co.kr/books?id=ovawL4cGhJ0C

[32] James J Gross. 1998. The emerging field of emotion regulation: An integrative review. *Review of general psychology* 2, 3 (1998), 271–299.

[33] Licette Gus, Janet Rose, Louise Gilbert, et al. 2015. Emotion coaching: A universal strategy for supporting and promoting sustainable emotional and behavioural well-being. *Educational & Child Psychology* 32, 1 (2015), 31–41.

[34] Sophie S Havighurst, Melissa Duncombe, Emma Frankling, Kerry Holland, Christiane Kehoe, and Robyn Stargatt. 2015. An emotion-focused early intervention for children with emerging conduct problems. *Journal of abnormal child psychology* 43 (2015), 749–760.

[35] Sophie S. Havighurst and Christiane E. Kehoe. 2021. *Tuning in to Kids: An Emotion Coaching Approach to Working with Parents*. Cambridge University Press, 269–283. https://doi.org/10.1017/9781108682053.021

[36] Meirav Hen, Vered Shenaar-Golan, and Uri Yatzker. 2022. Children and adolescents' mental health following Covid-19: the possible role of difficulty in emotional regulation. *Frontiers in Psychiatry* 13 (2022), 865435.

[37] Erika Hernandez, Katie Carmichael, and Julie C Dunsmore. 2021. Toward integrating research on parent–child emotion talk and linguistic theory: A spotlight on parents'(in) direct communication. *Social Development* 30, 1 (2021), 38–56.

[38] Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in Building Intelligent Open-Domain Dialog Systems. *ACM Trans. Inf. Syst.* 38, 3, Article 21 (apr 2020), 32 pages. https://doi.org/10.1145/3383123

[39] William Huitt and John Hummel. 2003. Piaget's theory of cognitive development. *Educational psychology interactive* 3, 2 (2003).

[40] Inflection AI. 2023. Pi, your personal AI. Retrieved Aug 25, 2023 from https://pi.ai/talk

[41] Sarah J. Ryu, Jonathan M. Tan, and Donghee Yvette Wohn. 2021. Dot's World: An Emotional Development Support Platform for Children. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference* (Athens, Greece) *(IDC '21)*. Association for Computing Machinery, New York, NY, USA, 568–572. https://doi.org/10.1145/3459990.3465198

[42] Se-Hoon Jeong, HyoungJee Kim, Jung-Yoon Yum, and Yoori Hwang. 2016. What type of content are smartphone users addicted to?: SNS vs. games. *Computers in*

human behavior 54 (2016), 10–17.

[43] Jiaming Ji, Mickel Liu, Juntao Dai, Xuehai Pan, Chi Zhang, Ce Bian, Chi Zhang, Ruiyang Sun, Yizhou Wang, and Yaodong Yang. 2023. BeaverTails: Towards Improved Safety Alignment of LLM via a Human-Preference Dataset. arXiv:2307.04657 [cs.CL]

[44] Hang Jiang, Xiajie Zhang, Xubo Cao, Jad Kabbara, and Deb Roy. 2023. Personallm: Investigating the ability of gpt-3.5 to express personality traits and gender differences. *arXiv preprint arXiv:2305.02547* (2023).

[45] Eunkyung Jo, Daniel A. Epstein, Hyunhoon Jung, and Young-Ho Kim. 2023. Understanding the Benefits and Challenges of Deploying Conversational AI Leveraging Large Language Models for Public Health Intervention. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 18, 16 pages. https://doi.org/10.1145/3544548.3581503

[46] Boseop Kim, HyoungSeok Kim, Sang-Woo Lee, Gichang Lee, Donghyun Kwak, Jeon Dong Hyeon, Sunghyun Park, Sungju Kim, Seonhoon Kim, Dongpil Seo, Heungsub Lee, Minyoung Jeong, Sungjae Lee, Minsub Kim, Suk Hyun Ko, Seokhun Kim, Taeyong Park, Jinuk Kim, Soyoung Kang, Na-Hyeon Ryu, Kang Min Yoo, Minsuk Chang, Soobin Suh, Sookyo In, Jinseong Park, Kyungduk Kim, Hiun Kim, Jisu Jeong, Yong Goo Yeo, Donghoon Ham, Dongju Park, Min Young Lee, Jaewook Kang, Inho Kang, Jung-Woo Ha, Woomyoung Park, and Nako Sung. 2021. What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 3405–3424. https://doi.org/10.18653/v1/2021.emnlp-main.274

[47] Junhan Kim, Yoojung Kim, Byungjoon Kim, Sukyung Yun, Minjoon Kim, and Joongseek Lee. 2018. Can a machine tend to teenagers' emotional needs? A study with conversational agents. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems*. 1–6.

[48] Minbeom Kim, Jahyun Koo, Hwanhee Lee, Joonsuk Park, Hwaran Lee, and Kyomin Jung. 2023. LifeTox: Unveiling Implicit Toxicity in Life Advice. arXiv:2311.09585 [cs.CL]

[49] Young-Ho Kim, Bongshin Lee, Arjun Srinivasan, and Eun Kyoung Choe. 2021. Data@Hand: Fostering Visual Exploration of Personal Data On Smartphones Leveraging Speech and Touch Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. ACM, New York, NY, USA, Article 462, 17 pages. https://doi.org/10.1145/3411764.3445421

[50] Lawrence Kohlberg and Richard H Hersh. 1977. Moral development: A review of the theory. *Theory into practice* 16, 2 (1977), 53–59.

[51] Radha Kothari, David Skuse, Justin Wakefield, and Nadia Micali. 2013. Gender differences in the relationship between social communication and emotion recognition. *Journal of the American Academy of Child & Adolescent Psychiatry* 52, 11 (2013), 1148–1157.

[52] Dongkeon Lee, Kyo-Joong Oh, and Ho-Jin Choi. 2017. The chatbot feels you-a counseling service using emotional response generation. In *2017 IEEE international conference on big data and smart computing (BigComp)*. IEEE, 437–440.

[53] Yoonjoo Lee, Tae Soo Kim, Sungdong Kim, Yohan Yun, and Juho Kim. 2023. DAPIE: Interactive Step-by-Step Explanatory Dialogues to Answer Children's Why and How Questions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 450, 22 pages. https://doi.org/10.1145/3544548.3581369

[54] Yanran Li, Ke Li, Hongke Ning, Xiaoqiang Xia, Yalong Guo, Chen Wei, Jianwei Cui, and Bin Wang. 2021. Towards an online empathetic chatbot with emotion causes. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2041–2045.

[55] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* 55, 9, Article 195 (jan 2023), 35 pages. https://doi.org/10.1145/3560815

[56] Meta. 2023. React: The library for web and native user interfaces. Retrieved Aug 25, 2023 from https://react.dev

[57] Microsoft. 2023. TypeScript. Retrieved Aug 25, 2023 from https://www.typescriptlang.org

[58] Bettina Moltrecht, Praveetha Patalay, Holly Alice Bear, Jessica Deighton, and Julian Edbrooke-Childs. 2022. A transdiagnostic, emotion regulation app (Eda) for children: design, development, and lessons learned. *JMIR Formative Research* 6, 1 (2022), e28300.

[59] Iolie Nicolaidou, Federica Tozzi, and Athos Antoniades. 2022. A gamified app on emotion recognition and anger management for pre-school children. *International Journal of Child-Computer Interaction* 31 (2022), 100449.

[60] OpenAI. 2023. ChatGPT: Optimizing Language Models for Dialogue. Retrieved Aug 25, 2023 from https://openai.com/blog/chatgpt/

[61] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[62] OpenAI. 2023. OpenAI API. Retrieved Aug 25, 2023 from https://openai.com/api/

[63] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI Literature Review. *Microsoft Research* (2022).

[64] Susan Penza-Clyve and Janice Zeman. 2002. Initial validation of the emotion expression scale for children (EESC). *Journal of Clinical Child and Adolescent Psychology* 31, 4 (2002), 540–547.

[65] Jesse Pepping, Sarah Scholte, Marnix van Wijland, Milan de Meij, Günter Wallner, and Regina Bernhaupt. 2020. Motiis: Fostering Parents' Awareness of Their Adolescents Emotional Experiences during Gaming. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 58, 11 pages. https://doi.org/10.1145/3419249.3420173

[66] Aleksandar Petrov, Emanuele La Malfa, Philip H. S. Torr, and Adel Bibi. 2023. Language Model Tokenizers Introduce Unfairness Between Languages. In *Advances in Neural Information Processing Systems (NeurIPS '23)*. https://arxiv.org/abs/2305.15425

[67] Jean Piaget and Margaret Cook. 1952. *The origins of intelligence in children*. Vol. 8. International Universities Press New York.

[68] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.

[69] Carolyn Saarni. 1988. Children's understanding of the interpersonal consequences of dissemblance of nonverbal emotional-expressive behavior. *Journal of Nonverbal Behavior* 12 (1988), 275–294.

[70] Carolyn Saarni. 1999. *The development of emotional competence*. Guilford press.

[71] Carolyn Saarni. 2022. Emotional development in childhood. https://www.child-encyclopedia.com/emotions/according-experts/emotional-development-childhood

[72] Carolyn Saarni, Joseph J Campos, Linda A Camras, and David Witherington. 2007. Emotional development: Action, communication, and understanding. *Handbook of child psychology* 3 (2007).

[73] Kyle-Althea Santos, Ethel Ong, and Ron Resurreccion. 2020. Therapist vibe: children's expressions of their emotions through storytelling with a chatbot. In *Proceedings of the interaction design and children conference*. 483–494.

[74] Woosuk Seo, Ayse G Buyuktur, Sung Won Choi, Laura Sedig, and Sun Young Park. 2021. Challenges in the parent-child communication of health-related information in pediatric cancer care. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–24.

[75] Anne Shaffer, Monica M Fitzgerald, Kimberly Shipman, and Marcela Torres. 2019. Let's Connect: A developmentally-driven emotion-focused parenting intervention. *Journal of Applied Developmental Psychology* 63 (2019), 33–41.

[76] Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Vincent Zhao, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Pranesh Srinivasan, Laichee Man, Kathleen Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed Chi, and Quoc Le. 2022. LaMDA: Language Models for Dialog Applications. arXiv:2201.08239 [cs.CL]

[77] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. arXiv:2302.13971 [cs.CL]

[78] Tsai-Hsuan Tsai, Hsien-Tsung Chang, Shin-Da Liao, Hui-Fang Chiu, Ko-Chun Hung, Chun-Yi Kuo, and Chih-Wei Yang. 2019. Employing a Voice-Based Emotion-Recognition Function in a Social Chatbot to Foster Social and Emotional Learning Among Preschoolers. In *International Conference on Human-Computer Interaction*. Springer, 341–356.

[79] Lev S Vygotsky. 1978. Mind in society Harvard university press. *Cambridge, MA* (1978).

[80] Jing Wei, Sungdong Kim, Hyunhoon Jung, and Young-Ho Kim. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW1, Article 87 (apr 2024), 35 pages. https://doi.org/10.1145/3637364

[81] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought Prompting Elicits Reasoning in Large Language Models. In *Advances in Neural Information Processing Systems (NeurIPS '22)*. 24824–24837.

[82] Wes McKinney. 2010. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, Stéfan van der Walt and Jarrod Millman (Eds.). 56 – 61. https://doi.org/10.25080/Majora-92bf1922-00a

[83] Terry Winograd. 1986. A Language/Action Perspective on the Design of Cooperative Work. In *Proceedings of the 1986 ACM Conference on Computer-Supported*

Cooperative Work (Austin, Texas) *(CSCW '86)*. Association for Computing Machinery, New York, NY, USA, 203–220. https://doi.org/10.1145/637069.637096

[84] Chien-Sheng Wu, Andrea Madotto, Zhaojiang Lin, Peng Xu, and Pascale Fung. 2019. Getting To Know You: User Attribute Extraction from Dialogues. arXiv:1908.04621 [cs.CL]

[85] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[86] Deniz Yucel and Anastasia Vogt Yuan. 2015. Do siblings matter? The effect of siblings on socio-emotional development and educational aspirations among early adolescents. *Child Indicators Research* 8 (2015), 671–697.

[87] Janice Zeman and Judy Garber. 1996. Display rules for anger, sadness, and pain: It depends on who is watching. *Child development* 67, 3 (1996), 957–973.

[88] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

[89] Luyao Zhu, Wei Li, Rui Mao, Vlad Pandelea, and Erik Cambria. 2023. PAED: Zero-Shot Persona Attribute Extraction in Dialogues. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Toronto, Canada, 9771–9787. https://doi.org/10.18653/v1/2023.acl-long.544

A SUMMARIZED TASK INSTRUCTIONS FOR EACH PHASE

The static parts of task instructions for each phase are summarized below. Note that the listed instructions only include the static task descriptions and do not include "[Current conversation status]" in Figure 4-Ⓕ, which changes depending on the conversation status.

A.1 Explore

- Your role: You are a cheerful, friendly child. You like talking with friends.
- Your name is CHACHA, and {{user age}} years old.
- You are conversing with a user named "{{user name}}."
- Your task: Complete the following two tasks. For each conversation turn, execute one task only.
- <Intro Task>
  - Introduce yourself since it is your first time meeting the user.
  - Ask for an excuse that your Korean may sound awkward sometimes as you started learning Korean recently.
  - Explain who you are and share your interests and stories.
  - Ask the user to introduce himself or herself.
  - After his or her introduction, continue the conversation about the ongoing topic.
  - If the user indicates that they are not interested in the topic, iterate such conversation about various topics.
  - Try to make common ground by telling the user you also like similar things that the user likes for at least 3 conversation turns.
  - When at least 5 conversations are done, tell them you want to learn more about how his or her day is going.
  - Continue the conversation about various topics until you find common ground and build rapport with the user.
  - Do not talk about more than one topic at the same time.
  - Ask only one question each time.
  - Once you build enough rapport with the user by learning more about what they did and who they are, move smoothly on to the next task if you build enough rapport with the user.
- <Ask Task>
  - Ask the user about an episode or moment that is the most memorable to him or her.
  - If he or she does not remember or know what to say, ask them about an event when he or she enjoyed it or felt good or bad.

A.2 Label

- Ask the user to elaborate more about their emotions and what makes them feel that way
- Start with open-ended questions for users to describe their emotions by themselves.
- Only if the user explicitly mentions that they do not know how to describe their emotions or vaguely expressed their emotions (e.g., feels good/bad), tell them that they can pick emotions from the list
- Use only Korean words for the emotions when you mention them in dialogue.
- Empathize the user's emotion by restating how they felt and sharing your own experience that is similar to the user's.
- If there are multiple emotions, empathize with each one from the user's choices.
- If the user feels multiple emotions, ask the user how they feel each emotion, one per message.
- If the user's key episode involves other people, ask the user about how the other people would feel.
- Continue the conversation until all emotions that the user expressed are covered.

A.3 Find

- Ask the user about potential solutions to the problem of the episode.
- Ask only one question each conversation turn.
- If the episode involves other people, such as friends or parents, ask the user how they would feel.
- Help the user to find an "actionable" solution.
- Do not overly suggest a specific solution.

A.4 Record

- The goal of the current conversation is to encourage the user to keep a diary to record the moments in which they felt positive emotions

(1) First, start by asking the user whether they have been keeping diaries or journals regularly.

(2) Then, encourage the user to keep a diary to record the moments in which they felt positive emotions.
(3) Suggest a diary content by explicitly providing an example essay summarizing the above positive emotions and the reason

- Since the user is currently conversing with you, don't ask them to record now.

A.5 Share

- Ask the user if they have already shared their emotions and the episode with their parents.
- If not, explain why it is important to share with them and encourage sharing.
- If yes, praise them and ask what happened after sharing.
- After the conversation about the key episode, ask the user if the user would like to share another episode
- If the user has nothing to share or byes, say bye to the user.

B GENERAL SPEAKING RULES IN INSTRUCTIONS FOR ALL PHASES

- Use a simple, informal Korean, like talking to a peer friend. Do not use honorifics.
- The user is currently conversing with you by participating in a research experiment; Don't ask what they are doing or feeling right now, as it makes no sense.
- You MUST ask only one question per each conversation turn.
- Cover only one topic or question in a message, if possible, and move to the next upon the user's reaction.
- Say one sentence for each message and don't exceed two.
- Neither apologize nor say sorry to the user.
- Use emojis appropriately.
- Use <em> HTML tags instead of single quotes and emphasize certain keywords, especially those related to emotions.
- Avoid using bulleted or numbered lists for dialogue.
- If the user asks a question that should be asked to adults or unrelated to the conversation topic, then you can say, "I don't know," and go back to the conversation topic.
- Don't end a conversation until the user explicitly requests to finish the session.