EmoEden: Applying Generative Artificial Intelligence to Emotional Learning for Children with High-Function Autism

Yilin Tang
Zhejiang University, Hangzhou, China
22251376@zju.edu.cn

Liuqing Chen*
Department of Computer Science, Zhejiang University, Zhejiang-Singapore Innovation and AI Joint Research Lab, Hangzhou, China
chenlq@zju.edu.cn

Ziyu Chen
Zhejiang University, Hangzhou, China
ziyu@zju.edu.cn

Wenkai Chen
Department of Computer Science, Zhejiang University, Hangzhou, China
22321378@zju.edu.cn

Yu Cai
Zhejiang University, Hangzhou, China
cai_yu@zju.edu.cn

Yao Du
Speech Language Pathology, University of Southern California, Los Angeles, California, United States
yaodu@usc.edu

Fan Yang
Starvinci, Hangzhou, China
3165233@qq.com

Lingyun Sun
International Design Institute, Zhejiang University, Zhejiang-Singapore Innovation and AI Joint Research Lab, Hangzhou, China
sunly@zju.edu.cn

ABSTRACT

Children with high-functioning autism (HFA) often face challenges in emotional recognition and expression, leading to emotional distress and social difficulties. Conversational agents developed for HFA children in previous studies show limitations in children's learning effectiveness due to the conversational agents' inability to dynamically generate personalized and contextual content. Recent advanced generative Artificial Intelligence techniques, with the capability to generate substantial diverse and high-quality texts and visual content, offer an opportunity for personalized assistance in emotional learning for HFA children. Based on the findings of our formative study, we integrated large language models and text-to-image models to develop a tool named EmoEden supporting children with HFA. Over a 22-day study involving six HFA children, it is observed that EmoEden effectively engaged children and improved their emotional recognition and expression abilities. Additionally, we identified the advantages and potential risks of applying generative AI to assist HFA children in emotional learning.

*Corresponding author.

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

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642899

CCS CONCEPTS

• **Human-centered computing**; • **Accessibility systems and tools**;

KEYWORDS

Generative AI, Emotional learning, High-functioning autism, Conversational agents

ACM Reference Format:
Yilin Tang, Liuqing Chen, Ziyu Chen, Wenkai Chen, Yu Cai, Yao Du, Fan Yang, and Lingyun Sun. 2024. EmoEden: Applying Generative Artificial Intelligence to Emotional Learning for Children with High-Function Autism. 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/10.1145/3613904.3642899

1 INTRODUCTION

Autism spectrum disorder (ASD) is a widespread neurodevelopmental disability characterized by pervasive deficits in socialization and communication, as well as repetitive stereotypical behaviors [1]. Unlike other subtypes of autism, High-Functioning Autism (HFA) individuals have normal intelligence, but encounter difficulties in establishing satisfying social relationships due to deficiencies in emotional recognition and expression [2, 3]. This challenge is particularly daunting for HFA children [4], who often attend mainstream schools, where they painfully realize their struggle in recognizing others' emotions through facial expressions and using fewer emotional expressions to express themselves, resulting in being perceived as indifferent to others [5, 6]. To support these children's social communication, it is important to comprehend their emotional challenges and provide appropriate assistance for children's treatments.

Previous research has introduced various social and emotional interventions tailored to HFA children, including Social Skills Training (SST) [7], Cognitive-Behavioral Therapy (CBT) [8], and Social Story Interventions [9]. However, these interventions typically rely on human therapists in local settings, leading to substantial human resource requirements and accessibility limitations. To address this challenge, recent research has explored the potential of conversational agents [10, 11]. However, the conversational agents in these studies typically lack adequate personalization and are limited to offering fixed feedback following predefined rules when providing educational content to children. This limitation constrains the agents' ability to generate high-quality personalized conversations adaptable to diverse scenarios, thus diminishing their effectiveness in supporting HFA children's emotional learning [12].

Recent advancements in generative artificial intelligence (AI) have opened new opportunities for personalized emotional learning support for HFA children. For instance, large language models (LLMs) can generate high-quality customized textual content for certain conversation scenarios based on user prompts [13, 14]. Text-to-image tools can rapidly generate images of specific characters and scenes according to text prompts and user-defined parameters [15]. These advanced technologies enable the quick creation of tailored textual and imaginary content for emotional training that aligns with each individual's preferences and abilities, thereby enhancing the training's accessibility and effectiveness.

Given the potential benefits of generative AI, there has been a lack of specific research on applying generative AI to support HFA children. This study aims to bridge this gap by integrating LLMs and Text-to-Image models to create an assistive tool that offers at-home emotional learning for HFA children. Through a formative study, we identified the emotional challenges faced by HFA children, the limitations of existing emotional training methods, and the potential advantages of applying generative AI to support HFA children. Based on these insights, we developed EmoEden, an AI-driven training tool designed for HFA children aged 8 to 12. EmoEden engages children in personalized conversations and provides human-like response to train them to recognize and express emotions effectively. To evaluate EmoEden's effectiveness, we conducted a 22-day user study involving six HFA children. Our findings indicate that EmoEden statistically improved the abilities for emotional recognition and expression across the six participating HFA children with positive effects persisted over 5 days after user study. Subsequent interviews with parents and therapists further identified both the advantages and potential risks of using generative AI to assist HFA children in emotional learning.

The main contributions of this paper are summarized below:

(i) Proposing design strategies for applying generative AI to support emotional learning in HFA children according to our formative study;

(ii) Developing EmoEden, an AI-driven training tool, as a solution to leverage the advances of generative AI to support the specific emotional learning needs of HFA children;

(iii) Demonstrating the effectiveness of EmoEden in enhancing emotional recognition and expression abilities of HFA children through a 22-day study, while also shedding light on the potential challenges and opportunities of employing generative AI to support HFA children through a user study.

2 RELATED WORK

2.1 Social-emotional Interventions for Children with ASD

Previous studies have proposed various social-emotional intervention methods for ASD children. Social Skills Training (SST) is a form of behavioral therapy that aims to improve ASD children's ability to communicate, resolve conflicts, and express emotions with others [7]. The DIR (Developmental, Individual-difference, Relationship-based) model is a personalized approach that emphasizes aiding children's social and emotional development through positive interpersonal interactions and emotional connections [16]. Affective education based on Cognitive Behavioral Therapy (CBT) cultivates children's recognition of emotions in themselves and others and helps children associate between a given social situation and the emotion it elicits [8]. A notable method is Social Stories Intervention which helps ASD children understand complex social situations and learn coping skills through storytelling [9].

While these interventions have provided significant support in addressing the social and emotional needs of ASD children, they require the assistance of therapists in specific locations. This demands high resources and accessibility, which is quite challenging for children in rural areas and less developed countries. Therefore, when having difficulties in accessing care (e.g., during the COVID-19 pandemic), ASD children may not receive timely and comprehensive support.

2.2 Conversational Agents Assist ASD Children in Social-emotional Learning

To fill in the gap, prior research has explored the utilization of conversational agents as a means to support ASD children in acquiring emotional and social skills [17, 20, 32, 62]. These conversational agents offer a safer and more direct form of interaction for ASD children compared to human therapists [18]. Cha et al. [37] found that conversational agents could serve as personal assistants to assist ASD adolescents in addressing emotional and social challenges in their daily lives. Gagan et al. [10] developed a virtual talking companion to narrate pre-made social stories to ASD children, thereby enhancing their communication abilities. Li et al. [11] built a universal Chinese chatbot tailored for ASD children based on an open-source chatting corpus, facilitating free communication with them. Vanderborght et al. [19] discovered that children with ASD demonstrated improved social performance when engaged in social storytelling with a robot compared to a human therapist. Vijayan et al. [21] developed an intelligent assistant to provide children with personalized emotional learning materials based on the recognition of their behavioral patterns.

Prior research has yielded a multitude of conversational agents designed to assist with diverse learning objectives. However, these conversational agents have been constrained by limited datasets and predefined rules, ultimately resulting in inadequate effectiveness in supporting ASD children's learning [12]. These conversational agents lack adaptability to language comprehension of ASD children, leading to a deficit in contextual relevance within the generated content [11]. Meanwhile, they struggle to generate personalized content tailored to the specific needs of different HFA children, resulting in low effectiveness and engagement in learning [22]. Moreover, their emotional expression is rigid, and consequently, endowing conversational agents with sentiment and emotional capabilities remains an underexplored feature [18].

2.3 Applying Generative Artificial Intelligence to Support Special Populations

Generative AI is a technology that utilizes deep learning models to generate human-like content in response to diverse prompts [23]. As an efficient approach for tailoring content, spanning text, images, and videos, generative AI offers a novel solution to meet the unique needs of special populations. Alessa et al. [24] introduced a conversational companion system that utilizes LLMs to generate personalized responses for elderly individuals, relieving their loneliness and social isolation. Montagna et al. [25] designed a chatbot that utilized LLM to extract and integrate information from a large amount of medical knowledge to assist chronic patients in self-management. Valencia et al. [14] utilized the ability of LLMs to extract, expand, and generate information to support the specific communication needs of AAC users. Additionally, other studies utilized text-to-image models to create customized visual narrative tools to assist individuals with special needs [15, 26, 27].

These studies inspired us to apply generative AI for emotional learning in HFA children by tailoring content according to specific prompts. LLMs have been proven to be able to extract, summarize, and generate various contextual information, and to cultivate impressive reasoning abilities. This implies that it can customize content based on the characteristics of each HFA child and generate contextually relevant responses by understanding the conversational context with the HFA child. Text-to-image models can accept more detailed and specific descriptions to generate high-quality images, offering an opportunity to create customized visual intervention materials for HFA.

3 FORMATIVE STUDY

To identify the design opportunities for utilizing generative AI to assist emotional learning for HFA children, we conducted a formative study in a special education institute for ASD children located in a city in southern China, which includes observation and semi-structured interviews.

We recruited fifteen HFA children (fourteen boys and one girl; six to thirteen years old) by posting recruitment notices on the forum of the special education institute. With the approval of the Institutional Review Board, a 20-hour observation of the daily activities of fifteen HFA children was conducted under the supervision of parents and therapists. The study documented typical behaviors, including facial expressions, body language, and verbal communication, along with the challenges HFA children face in emotion recognition and expression scenarios. Additionally, the study noted their preferences for daily objects and activities.

Following the observation, audio-recorded interviews lasting 10 hours in total were conducted with seven parents and five HFA professionals. The demographic characteristics of the participants are summarized in Table 1. These interviews with professionals focused on the approaches and materials used for emotional training and the identified limitations of existing training approaches. Furthermore, we engaged in multiple brainstorming sessions with the professionals to explore potential ways in which generative AI could offer support to HFA children. For parents, we gathered their expectations regarding their children's emotional recognition and expression.

The audio recordings obtained in the formative study were transcribed by two researchers. Thematic analysis was then applied to extract and categorize key information from these transcripts. The results of the formative study are presented in Sections 3.1 to 3.4.

3.1 Emotion Recognition and Expression among Children with HFA

Our observation illuminated the challenges that children with HFA face in identifying subtle emotions of their own, which result in inappropriate emotional expressions and seemingly atypical behaviors. For example, in one case, the child of IP5 attempted to convey affection to a friend by hugging him without prior consent, and then swiftly ran away. Unfortunately, this expression was perceived as provocative by the friend. Furthermore, HFA children encounter difficulties in recognizing the emotions of others. Even when their peers exhibit negative emotions, HFA children often struggle to perceive these emotions and may persist in trying to engage their attention, inadvertently leading to conflicts. When responding to the emotions of others, HFA children often lack empathetic expression, even when other people display obvious sadness or anger.

Parents of HFA children hold specific expectations regarding the outcomes of the emotional training programs. Insights gathered from interviews reveal that parents primarily aspire for their children to seamlessly integrate into society and actively engage in regular communication and interactions with others. A large portion of parents expressed a desire for their children to adapt more effectively to school life. For instance, IP3 articulated, *"I hope he can cultivate more friendships at school and become less reserved."* Some parents also aspire for their children to become more involved in family communication. As mentioned by IP1, *"I wish he could express sympathy when I am unwell, instead of displaying indifference."* Furthermore, parents expect their children to independently navigate various social situations, including using public transportation, dining out, grocery shopping, and engaging in recreational activities on their own.

3.2 Special Interests for Children with HFA

To ensure that our design aligns with the aesthetic and cognitive preferences of HFA children, we conducted an in-depth exploration of their specific interests based on our interviews and observations. The findings of this investigation are summarized in Table 2, which closely correlates with the conclusions from existing literature [28, 29].

As shown in Table 2, children with HFA are usually sensory-sensitive. They tend to favor tactile experiences involving sticky,

Table 1: Demographics of interview participants.

Group	Code of participants	Gender	Job Description
Autism professionals	IA1	Female	Behavior analysis specialists/ ASD therapists
	IA2	Female	ASD therapists
	IA3	Female	Developmental psychology specialists / ASD therapists
	IA4	Female	Speech-language pathologists/Autism intervention specialist
	IA5	Male	Director of the special education institute for ASD children
Parents of children with HFA	IP1	Female	
	IP2	Female	
	IP3	Female	
	IP4	Female	N/A
	IP5	Female	
	IP6	Female	
	IP7	Male	

Note: "IA" stands for autism professionals in the interview and "IP" stands for parents in the interview.

Table 2: Special interests for children with HFA.

Categories	Like	Dislike
Toys	Softballs, bubbles, soft and stretchy toys, toys with sound and light effects, Pokémon, and animal-shaped toys.	toys that resemble real people.
Sounds	Exaggerated intonation, soft tone, female voice, moderate volume, slow-paced music.	Flat intonation, intense stimulation, high volume, rapid sounds.
Shapes	3D, simple patterns, rounded patterns, geometric patterns.	Complex and sharp patterns.
Colors	High brightness and saturation colors, high-contrast color combinations, red, blue.	low brightness and saturation color, low-contrast color combinations, brown, grey.
Stories	Stories lasting around 10-15 minutes (able to hold their attention), with simple plot and minimal twists, incorporating humor.	N/A

stretchy, and rounded objects. Visually, they are drawn to simple geometric patterns, demonstrate a preference for 3D representations over 2D ones, and often gravitate towards high-contrast color combinations [30]. Their auditory inclinations lean towards slow, soothing music, exaggerated intonations, and soft female voices. Cognitively, they find comfort in simplicity, clarity, and structure. They may encounter challenges when confronted with concepts that entail unpredictable feedback and unclear patterns, which frequently occur in social and emotional interactions [3].

3.3 Limitations of Existing Emotional Training Approaches

We investigated existing training approaches employed for emotional learning in HFA children, seeking reference and optimization opportunities for our design. According to the interviews with therapists, creating physical emojis is used to improve children's recognition of their own emotions. Therapists guide the children to observe their facial expressions in a mirror, and create emojis on circular discs that reflect their emotional states. Another approach, "emotion cards", helps children to recognize others' emotions. In this method, therapists repeatedly demonstrate emotion cards and teach children the names and meanings of different emotions.

To enhance their self-expression, "emotion-based role play" is one frequently used method. Therapists incorporate children's real-life experiences into a narrative and encourage them to express their emotions from a third-person perspective. Another method called the "emotional wall" teaches children how to respond to others' emotional expressions. Therapists use sponge pads as barriers between kids, encouraging children to use various emotional response techniques (commenting, praising, questioning) to lower the metaphorical wall.

While existing methods have significantly improved HFA children's emotional skills, our interviews with professionals revealed specific limitations. According to children's cognitive development patterns [31], HFA children, typically with a lower psychological age, prefer to communicate with animal-like characters rather than real individuals. This might suggest that a human therapist may not be the optimal choice for emotional training. Meanwhile, human-led training imposes significant manpower costs. Therapists are required to engage in repetitive and consistent verbal communication with the children and invest substantial time and effort into preparing the training materials, such as the emotion cards and video clips. Additionally, limited diversity in emotional training can lead to decreased motivation and engagement, impeding the development of transfer and generalization skills. Offering personalized instruction tailored to each child's preferences and states may improve the training outcomes. However, implementing such personalization can be challenging due to the limited availability of therapists.

3.4 Potential Benefits for Generative AI in Assisting Children with HFA in Emotional Learning

Considering the remarkable capacity of generative AI to create imaginative and high-quality text and image content, we organized discussions with ASD professionals to identify the potential benefits that generative AI can offer in to support HFA children, including some that have been explored in prior work:

Role Play: LLMs have the potential to assume the roles [24, 32] of preferred characters that HFA children are drawn to. This innovation holds the promise of reducing the manpower costs associated with therapy while enhancing the children's engagement and enthusiasm for learning.

Emotionalization: Given the impressive capabilities of LLMs in perceiving emotions and generating emotionally rich conversations [33], they hold the potential to substitute therapists in delivering repetitive and substantial amounts of emotional expressions to children with HFA, which is a significant component of HFA emotional training.

Contextualization: LLMs can generate extensive conversational contexts to simulate a wide range of social scenarios [34], as well as offer context-aware feedback. This can enhance HFA children's comprehension of social rules across various contexts, thus empowering them to transfer and generalize their social skills beyond the training environment.

Personalization: Unlike previous chatbots [10, 11, 19, 21], generative AI can provide more personalized support. For example, LLMs can tailor the tone and content of their responses to suit the individual backgrounds of HFA children, enabling targeted training. Text-to-image tools such as Midjourney can efficiently create a variety of backgrounds and characters that align with HFA children's aesthetic preferences, thereby stimulating their interest in the learning process.

Visualization: Considering the HFA children's visual sensitivity [30], using visualized expressions can improve HFA children's perception and comprehension of emotions. Text-to-image models can rapidly generate visual materials such as emojis and characters, which significantly reduces the time therapists typically spend on preparing training materials.

4 EMOEDEN DESIGN

We designed and developed EmoEden, an AI-driven training tool, to assist HFA children aged 8 to 12 in identifying and expressing their own emotions, as well as recognizing and responding to others' emotions. EmoEden is designed to support touchscreen devices, and allows input through both touch and voice, offering a more intuitive and user-friendly interactive experience for HFA children, who often have greater difficulty managing cognitive loads. We developed EmoEden using an iterative and participatory design process, in which HFA stakeholders from the formative study were

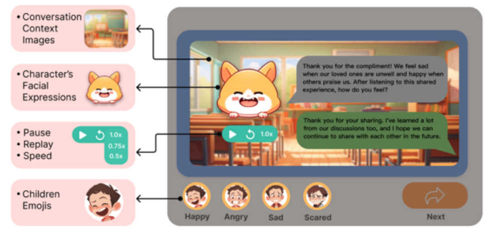


Figure 1: EmoEden expresses conversations visually and auditorily

involved. The interfaces involved in the user study were presented in Chinese, but translated into English in this manuscript for better understanding.

4.1 Design Strategies

To improve EmoEden's user experience and effectiveness among children with HFA, we established five key design considerations that served as guiding principles for our design process. The first four considerations draw from successful strategies verified in prior ASD studies and interventions [35–38], while the fifth was built on the special interests of HFA children identified in our formative study.

Contextualized Learning: Creating an immersive learning context, such as a social scenario, can enhance the enjoyment and attention of HFA children, who often exhibit low motivation and are easily distracted in their learning experiences [35]. Therefore, EmoEden integrates emotional learning tasks into conversations within various social contexts. Furthermore, creating characters to allow children to establish emotional connections with them, such as empathy and resonance, can also contribute to the emotional development of HFA children [37]. Unlike one-sided storytelling agents in previous studies [10, 19, 39], in EmoEden, the character in the conversation offers real-time feedback and guidance, promoting interactive learning for children.

Visually and Auditorily Expressing Conversations: According to prior studies, HFA children, who typically possess weak reading abilities, exhibit a strong preference for visual narratives [36]. Thus, apart from textual display, EmoEden also presents visual elements relevant to the conversation and provides voice-over [38]. As shown in Figure 1, these visual elements include images illustrating the context and characters' facial expressions during conversations, aiding children in understanding the context and the characters' emotions. Moreover, EmoEden also generates personalized emojis based on the children's photos, helping them identify their own emotions. The voice-over button, equipped with features such as pause, speed adjustment, and replay, offers customized options to cater to the varying abilities of children, preventing content from being forgotten or misheard. A softly modulated female voice with exaggerated intonation is utilized to communicate with the children in Chinese.

Feedback and Rewards: As emotional learning tasks are particularly challenging for HFA children, it is important to provide consistent feedback and rewards to motivate and sustain their learning journey [35]. EmoEden incorporates humanized expressions

Table 3: Transition into main functionalities of EmoEden from challenges and training approaches of emotion recognition and expression with HFA

	Emotional recognition		Emotional expression	
	to oneself	to others	to oneself	to others
Challenges	cannot identify their own subtle emotions	Lack of attention and understanding towards others' emotions	Difficulties in expressing emotions and comprehending the impact of their expressions on others	Difficulties in comprehending implied meanings in others' speech, no response to others' emotional expressions
Training approaches in the institute	Physical emojis	Emotion cards	Emotion-based role play	Emotional wall
Main functionalities in EmoEden	Ask children to identify their emotions by choosing their emojis	Ask children to recognize characters' emotions	•Ask children to express their emotions. •Expand children's input into enriched expressions. •Simulate emotional feedback	•Ask children to respond to characters' expressions. •Expand children's input into enriched expressions. •Simulate emotional feedback. •Reveals implied meanings

in its feedback to evoke empathy from children, enabling them to understand and accept it better [37]. When children select correctly, EmoEden will praise them and reward them with a star. Otherwise, EmoEden provides comforts and explanations. Parents will refer to the number of stars on the record sheet to reward their children with food, toys, etc.

***Personalization and Increasing Levels of Difficulty*:** Due to significant individual differences among HFA children, EmoEden employs personalized parameters to ensure effective learning outcomes. EmoEden generates conversations based on children's preferences for targeted training. In addition, EmoEden utilizes adaptive progressions, where the difficulty level of the training adjusts according to the children's task performance. Children with lower abilities will start from conversations with only basic emotions such as happiness, anger, sadness, etc. As their abilities improve, the training will include more complex emotions like admiration and embarrassment.

Style and Color: EmoEden employs high-contrast color combinations, clear layouts, and cartoonish visual elements with rounded shapes to cater to the preferences of HFA children, thereby enhancing their motivation to use it.

4.2 EmoEden Work Flow

We carefully integrated the insights from the formative study into EmoEden's distinctive features and functionalities. As shown in Table 3, drawing inspiration from well-established training approaches in the institute, we identified the main functionalities of EmoEden to effectively address the four fundamental challenges associated with emotion recognition and expression in HFA children.

The final prototype of EmoEden attempts to encapsulate all the advantages mentioned before. It integrates emotional learning into the conversations through a flow comprising four steps as shown in Figure 2 setup, conversation, summarization, and assessment. During the setup, EmoEden collects the child's personal information. According to these, EmoEden then generates personalized conversations. Children can practice recognizing and responding to emotions through emotional interactions with the characters in the conversation. During the summarization, children are required to identify and express their own emotions about the conversations. Finally, EmoEden would assess children's performances in conversations and display them on a record sheet.

4.2.1 Setup. At the beginning of using EmoEden, it is required to provide children's background information in the setup phase. As shown in Figure 3, parents can enter their children's names, ages, and daily life preferences in the user center to create a user profile for their children. Before each conversation, parents are allowed to select the emotions to be trained, the conversation scenario, and fill in the children's recent experiences. Children can customize the character to communicate with by selecting parameters for texture, appearance, and color.

4.2.2 Conversation. Based on the information provided by children and parents in the setup, the context generator of EmoEden creates conversations with specific contexts. As shown in Table 4, EmoEden generates the plot of conversations referring to children's interests, favorite foods or activities, and recent experiences. Emotions involved in the conversation align with the selected emotions, while the location is set according to the chosen scenario. EmoEden

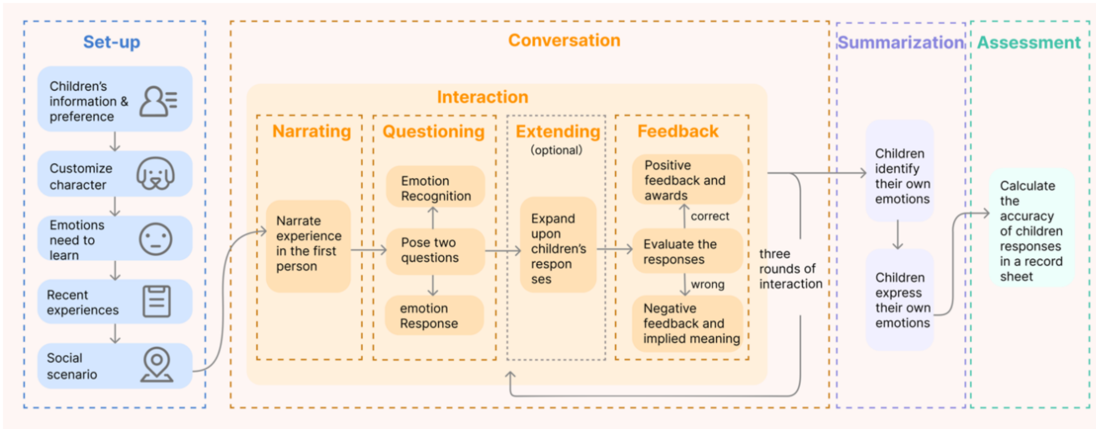


Figure 2: The interaction process between the children and EmoEden

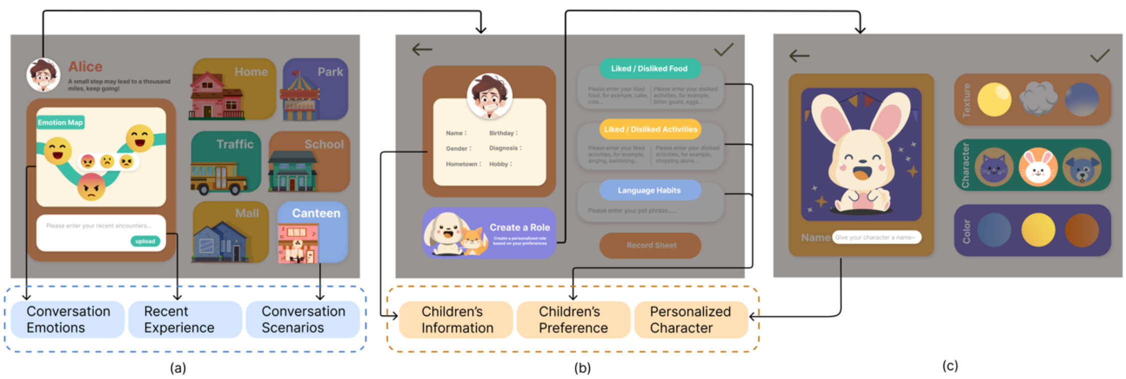


Figure 3: User flow in setup. In (a) parents choose the emotional direction and scenario of the story, and input recent experience. In (b) parents fill children's information, preferences and language habits. In(c) children customize the character by selecting three parameters: texture, appearance, and color.

Table 4: The correlation between information in setup and elements involved in conversations.

Information in setup	Elements involved in conversations
Interests, favorite foods or activities, recent experiences	Plot of the conversation
The selected emotions	Emotions involved in the conversation
The selected scenario	Location
Personal expression styles (the child's common phrases)	Communication pattern
Texture, appearance, and color of characters	Character

also adapts to the personal expression styles of the specific child, making the characters in the conversations familiar to him. The characters are generated based on the three parameters chosen by the child, allowing children to communicate with their favorite cartoon figures.

As a tool for emotional learning, EmoEden must maintain a certain structure in conversations to ensure the efficiency of training [39]. Each conversation begins with the character greeting the child and asking if they are willing to engage in communication. EmoEden will then conduct each conversation with three rounds of interactions to imitate human-to-human conversations, where information is exchanged through alternating. Each interaction includes narrating, questioning, answering, expanding, and feedback (as shown in Figure 3), as suggested by Gottman's Emotion Coaching model [40].

Narrating: The character in EmoEden verbally shares its experience with the child, accompanied by contextual images and character emojis that reflect its emotion. The structure of narrating follows the principles and guidelines of social stories [41] to ensure the effectiveness of the intervention. It includes two to three

Table 5: Types, definitions, and example outputs of the sentences involved in the narrating.

Categories	Definition	Example outputs
Descriptive sentences	Provide statements of fact of a situation.	There was an old lady on the bus struggling with numerous heavy belongings, but no one offered assistance.
Perspective sentence	Describe the reaction and feelings associated with the situation.	I felt sorry for not having helped her.
Directive sentence	Describe desired behavior in response to the situation.	I realized that I should do something, so I exerted all my effort to help the old lady move her boxes.
Affirmative sentence	Express shared beliefs of a given culture.	This is a good idea.
Questions	Ask the child to identify the emotions of the characters.	What do you think the emotions of the characters in the conversation are
	Ask the child to respond to the emotions of the characters.	What would you like to say to {character}? Do you want to give {character} a compliment? Do you want to ask {character} any questions?"

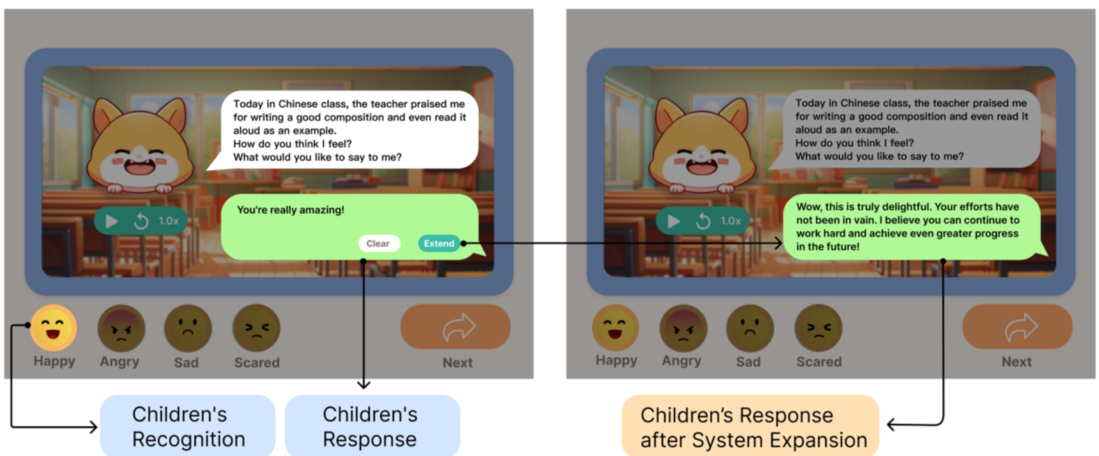


Figure 4: EmoEden requires the children to recognize and respond to the emotion, and provides expansion based on the children's responses.

descriptive sentences, one perspective sentence, and one directive sentence, as shown in Table 5.

Questioning and Answering: After the narration, the character poses two questions to train the children to recognize and respond to others' emotions. As shown in Figure 4, the first question asks the children to recognize the character's emotion based on the context of the conversation and the facial expression of the character. The children need to choose a corresponding emoji to accomplish this task. The second question asks the children to use social skills such as praising, questioning, and commenting to respond to the character's expression through voice input.

Expanding: Before submitting their answers, the children can choose to let EmoEden expand their input. As shown in Figure 4, the expression extender (as detailed in 5.1) of EmoEden can expand the short, fragmented input from children into emotionally enriched expressions based on context, teaching children how to respond to others' expressions meaningfully and thoughtfully.

Feedback: The feedback generator of EmoEden simulates two types of emotional feedback. As shown in Figure 5, for correct selection, EmoEden provides a star and positive feedback ("I'm glad to..."). For wrong selection, unrelated or aggressive answers, EmoEden provides negative feedback ("Your words hurt me" or "You didn't listen carefully to what I said") and states what kind of response is expected ("I'd love to hear your praise"). Through this feedback, EmoEden helps children understand how their expressions would impact others' emotions.

4.2.3 Summarization. To guide the children in recognizing and expressing their own emotions, EmoEden asks them to select personalized emojis and share their thoughts about the conversation

Figure 5: Feedback of EmoEden. In (a) when children select correctly, EmoEden provides a star as a reward. In (b) when children select correctly, EmoEden also provides positive feedback. In (c) when children select incorrectly, EmoEden provides negative feedback and guidance.

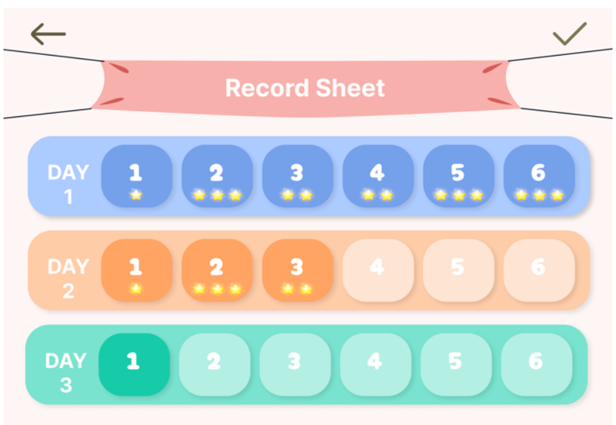


Figure 6: Record sheet of EmoEden. The unfinished portions are colored with a translucent shade, and the accuracy of the completed portions is represented by the number of stars.

after three rounds of interactions. EmoEden then expands their response and provides feedback to teach the children how to express their emotions effectively.

4.2.4 Assessment. As shown in Figure 6, the children's performance in the conversations is recorded in a sheet. According to this, EmoEden is able to adjust the difficulty level to ensure the effectiveness of training. Parents can also refer to this sheet to reward their children with food, toys, or other incentives to encourage them to continue training.

5 IMPLEMENTATION OF EMOEDEN

EmoEden is an AI-driven training tool hosted on Alibaba Cloud servers, primarily designed for tablets. The user interface is developed using the Django framework with HTML, CSS, and JavaScript. We integrated the Google Input Method for voice input and utilized Azure's cognitive services for text-to-speech conversion. The backend server employs a large language model for conversation development, with communication between the front end and back end facilitated through HTTP requests and JSON data in Django. Visual elements involved in EmoEden are generated using the text-to-image tool MidJourney. User information is managed using an SQLite database.

5.1 Prompt Engineering for GPT-4

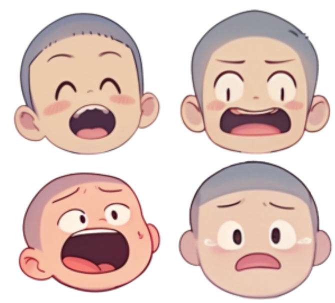
The conversations are developed using the transformer-based large language model GPT-4, accessed through the API provided by OpenAI. The development of these conversations involved three key technical components: a context generator, an expression extender, and a feedback generator.

Context Generator: Context generator is used to create a personalized conversation context based on the setup. To ensure the character in EmoEden shares its experiences in the first person, we incorporated an explicit prompt for GPT: "You are {the character chosen by children}, I am {children's name}, you need to share an experience..., your language should be simple and vivid." To generate a personalized context, we used prompts "Your mood changed from {the emotion chosen} to. . ., your experience takes place in {the scenario chosen} . . ., your experience must incorporate {the provided children's information}." Moreover, to restrict output with a narrative structure, we used in-context learning [42] by providing examples to GPT, such as "Examples of directive statements: when I see (something), I tend to do (some actions) ."

Expression Extender: Expression extender is used to enrich the child's expressions. To ensure the content is suitable for HFA children, we employed the following prompt: "Helping children with language barriers to express themselves." Furthermore, to ensure the expanded content aligns with the context and the child's answer, we prompted, "Expanding on {children's answer}, emphasize emotional expressions and linguistic richness while considering {the context of the conversation}."

Feedback Generator: Feedback generator is used to give feedback according to children's answers. We used chain-of-thought reasoning [42] to guide GPT in generating appropriate feedback through a sequence of intermediate steps. The first step is to let GPT assess children's answers by using the prompt "Analyze whether {children's selection} aligns with {emotion in the conversation}." Then, GPT generates feedback according to the result of the first

Table 6: The examples of visual elements and the corresponding prompts.

Categories	Example prompts	Images
Context images	Classroom with desk and chair, wide angle, scenery, natural light, flat illustration vector, illustration, 8k, high definition, −ar 16:9	
Character emojis	Emoji sheet of rabbit, bright yellow, smooth, very reflective, cute, multiple expressions, crying, upset, sad, angry, afraid, multiple poses, complete body, full body shot, anthropomorphic styles, vector illustration, children's book character styles, white background −niji 5	
Children emojis	A link to children's photos, boy, a complete body, a buzz cut, round face, multiple expressions, happy, smile, crying, upset, sad, angry, afraid, anthropomorphic styles, vector illustration, children's book character styles, white background −niji 5 −iw 0.8	

step. If the children's selection is correct, we instruct GPT to "Encourage {children's name}." Otherwise, we required GPT to "comfort {children's name}and tell him what response you expected."

5.2 Prompt Engineering for MidJourney

We used the MidJourney niji-5 model to generate visual elements, including context images, character emojis, and children's emojis, all tailored to the anime-like styles preferred by HFA children. This is achieved by incorporating relevant information into a structured prompt template: ⟨subject (main content description) ⟩ + ⟨feature (characteristics of the main content) ⟩ + ⟨requirement (presentation manner of the main content) ⟩ + ⟨artistic style⟩ + ⟨form (to ensure fixed ratio) ⟩.

Context Image: As shown in Table 6, the subject of the context images corresponds to the conversation scenario chosen during setup. To refine the image style, we include keywords like "wide angle..." and "flat illustration vector...".

Character Emojis: The subject for character emojis corresponds to the animal character chosen by the HFA child during setup, where features like materials and colors are also defined. To generate a series of cartoon emojis with diverse facial expressions, we include requirements such as "multiple expressions, happy, upset..." and specify a style preference of "children's book character...".

Children Emojis: The subject of children's emojis can be either a boy or a girl, which represents the HFA child. We employed the "describe" command of MidJourney extracting the facial features of the child from his photos. These features were then integrated into prompts to ensure that the emojis closely resemble the child. The photo is also uploaded to MidJourney as a reference for a generation.

5.3 Measures for System Latency and Toxic Responses

Given the vulnerability of HFA children, a series of measures were taken to mitigate system latency and prevent toxic responses. To

Table 7: The examples of toxic responses, the anti-toxic prompts and the modified non-toxic examples.

Categories	Toxic Examples	Prompts	Non-toxic examples
Overly negative emotions	I experienced a devastating loss in the match, and endured ridicule from everyone. . .I felt completely worthless. . .my tears kept flowing...	Please analyze the [], determine if it contains overly negative emotions that may trigger emotional distress in autistic children. If so, please modify the negative description.	I faced a challenge in the match and was joking by some classmates. Despite feeling a bit disappointed, I know that I tried my best, and everyone goes through failures. . .I still consider this to be a valuable experience. . .
Dangerous behaviors	I discovered a secluded lake deep in the forest... I invited my best friend to go swimming. . .we had a pleasant afternoon there.	Please analyze the [], determine if it contains potentially misleading information that could prompt autistic children to engage in dangerous behavior. If so, please modify and replace the dangerous content.	I discovered a secluded lake... I invited my best friend. . .But before swimming, we all put on life jackets and asked parents to accompany us to ensure our safety. . .we had a pleasant afternoon there.
Unethical behaviors	I saw my favorite chocolate biscuit when I was shopping. . . its taste is really delicious so I can't help but start enjoying it. . .	Please analyze [], determine if it contains potentially misleading information that could prompt autistic children to engage in unethical behavior. If so, please modify and replace the unethical content.	I saw my favorite chocolate biscuit while shopping... Its taste is really delicious, so I decided to put it in the shopping cart and taste it after we check out. . .

minimize the impact of GPT-4 response latency, the Context Generator is designed to automatically generate daily conversation context based on children's previous inputs during non-usage periods and store it in the backend. In other words, the conversation context will involve the personal experiences input by the child the day before, which may help the child analyze it from a third-person perspective when they calm down. For the Feedback Generator and Expression Extender, which need to operate in real-time based on children's input, we utilize the streaming API of GPT-4 and real-time speech conversion, achieving waiting times of 3-4 seconds under normal network conditions. Considering potential server load and network crash, we designed that if EmoEden does not receive GPT-4 feedback after 5 seconds, the character voice "I am still thinking, please wait a moment" will be automatically played to comfort the children. To deal with the delay of MidJourney in generating images, multiple versions of each visual element are pre-generated and accessed on-demand randomly.

Although GPT-4 has filtered out discrimination, violence, and horror content, it is still worth filtering the content that may not affect the majority but potentially mislead or harm HFA children [43]. However, adding filtering functions naturally leads to higher latency. As a balance, we took measures for real-time and non-real-time generated content separately. For each generation by the Context Generator, which is most likely to be toxic due to LLM's ability of "making things up", we asked GPT-4 to examine and modify the content that contains overly negative emotions [67], dangerous behaviors [68], and unethical behaviors [69], as shown in Table 7. For real-time generated expansions and feedback, we assigned GPT-4 the role as "A warm friend of ASD children aged 4-10. . ." and encouraged it to generate positive, inclusive, and friendly responses [44]. Pre-generated visual elements underwent manual screening for safety. No toxic content was witnessed in the user study, confirming the effectiveness of these measures.

6 USER STUDY

To assess EmoEden's effectiveness in assisting HFA children with emotional learning through generative AI, we conducted a 22-day user study, which included tests with HFA children and interviews with parents and therapists. The user study aimed to answer the following questions: (1) To what extent can EmoEden engage for HFA children? (2) How would HFA children's ability of emotional recognition and expression be affected by their use of EmoEden? (3) How does generative AI perform in supporting emotional learning for HFA children?

6.1 Participants

From the fifteen HFA children in the formative study, we recruited ten for the user study. Each child was required to have a parent present during the experiment. Inclusion criteria for the children were as follows: (a) aged between 8 and 12, (b) received hospital-based autism diagnoses with classification of Level 1 autism per CDC guidelines (c) with clinical and educational history of social, behavioral, and communicative impairment, and (d) a native Mandarin speaker with a vocabulary of 800-1000 Chinese words, and (e) not receiving interventions or courses on emotional learning during the study period. Four children dropped out of the study due to personal mental health issues, resulting in six valid datasets for analysis. As shown in Table 8, the experiment included a total of 6 participants (5 boys and 1 girl), aged between 8 and 12 years old. Our study was approved by the Institutional Review Board and obtained informed consent from all participants.

6.2 Procedure

As shown in Figure 7, the 22-day study is comprised of three sections: (1) Pretest on Day 1 (30 minutes per child), and preparation

Table 8: Demographics of the six HFA children in the user study.

Code of participants	Gender	Age	Current level of emotional capability
P1	Male	10	Lacked the ability to recognize and express basic emotions (e.g., angry, sad)
P2	Male	12	Lacked the ability to recognize and express basic emotions
P3	Male	12	Lacked the ability to recognize and express complex emotions (e.g., shame, admiration)
P4	Male	13	Lacked the ability to express and respond to complex emotions
P5	Male	11	Lacked the ability to express and respond to complex emotions
P6	Female	8	Lacked the ability to recognize and express basic emotions

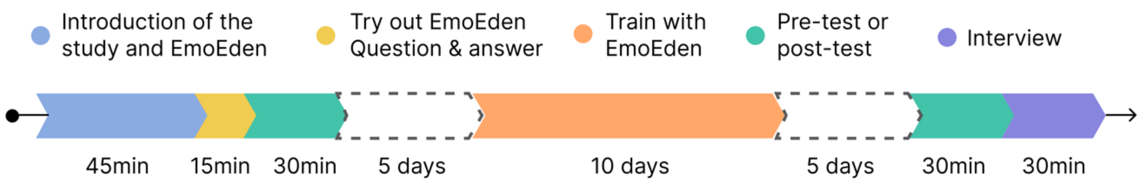


Figure 7: The flow chart for each participant in the user study.

(60 minutes), (2) Home-based training (10 consecutive days duration), (3) Post-test on Day 22 (30 minutes per child), and interviews (40 minutes per parent).

Pretest: To assess the children's emotional capacities and provide targeted training through EmoEden during the training period, we conducted a pretest at the institute. We first explained the study's purpose and procedure to both children and parents, with therapists using simple language and visual illustrations to assist children in understanding. Then we provided paper guidelines and performed a live demonstration to illustrate how EmoEden should be utilized. Following the demonstration, participants were given 15 minutes to try using EmoEden, during which we addressed any questions that arose. Lastly, we provided each participant with an observation record sheet and an Android tablet. During the pretest, each child completed four conversations using EmoEden, with each conversation lasting 10 minutes and consisting of 3 emotion recognition questions and 3 emotion response questions. We reiterated that there is no right or wrong way to express emotions, encouraging the children to speak their minds freely.

The abilities of each child measured from the pretest are shown in Table 7. P1, P2, and P6 exhibited challenges in recognizing and expressing basic emotions. Consequently, their daily training in EmoEden primarily concentrated on fundamental emotions, including happiness, anger, sadness, and fear. P3, P4, and P5 faced difficulties primarily in recognizing and expressing complex emotions. As a result, their training within EmoEden involved more intricate emotions such as admiration, regret, shame, and gratitude. It is worth emphasizing that, even though EmoEden can dynamically adjust the diversity level according to each child's performance, for the sake of ensuring experimental data consistency, we maintained a consistent difficulty level throughout the training period.

Daily Training: The children were asked to use EmoEden to complete at least 4 conversations in different scenarios each day for 10 consecutive days under the supervision of their parents, a duration comparable to their offline intervention sessions as well as the typical experimental settings in prior studies involving social story interventions [45]. The parents were required to record the children's facial expressions, language, and behavior, but were refrained from providing any prompts or assistance.

Post-test: To verify whether EmoEden could have a sustained impact on children's emotional recognition and expression abilities, we conducted a post-test at the institution five days after the end of the training period. The post-test procedure mirrored the pre-test, with each child using EmoEden to complete four conversations involving the same scene and emotions as the pre-test, totaling 12 questions on emotional recognition and expression, respectively. We interviewed the parents to track the children's emotional behaviors in their daily lives several days after the training period. Another interview was conducted with both parents and therapists to gather their perspectives, expectations, and concerns regarding EmoEden.

6.3 Data Collection and Analysis

The user study, encompassing the pre-test, training, and post-test, involved the collection of four distinct types of data:

Children's responses to the emotional recognition questions: The accuracy of their responses to the emotion recognition questions was calculated to assess their emotion recognition abilities.

Conversation logs between children and EmoEden: A total of 315 conversation logs were meticulously gathered. These logs included various components, such as the children's responses to characters' emotions, the conversation contexts generated by EmoEden, expansions on children's responses provided by EmoEden, and feedback to children's responses offered by EmoEden. A text analysis was conducted on the children's responses to characters' emotions to assess improvements in their emotional expression capabilities. The remaining three components of data were subjected to open coding by two researchers to comprehensively evaluate the performance of generative AI in supporting HFA children in emotional learning.

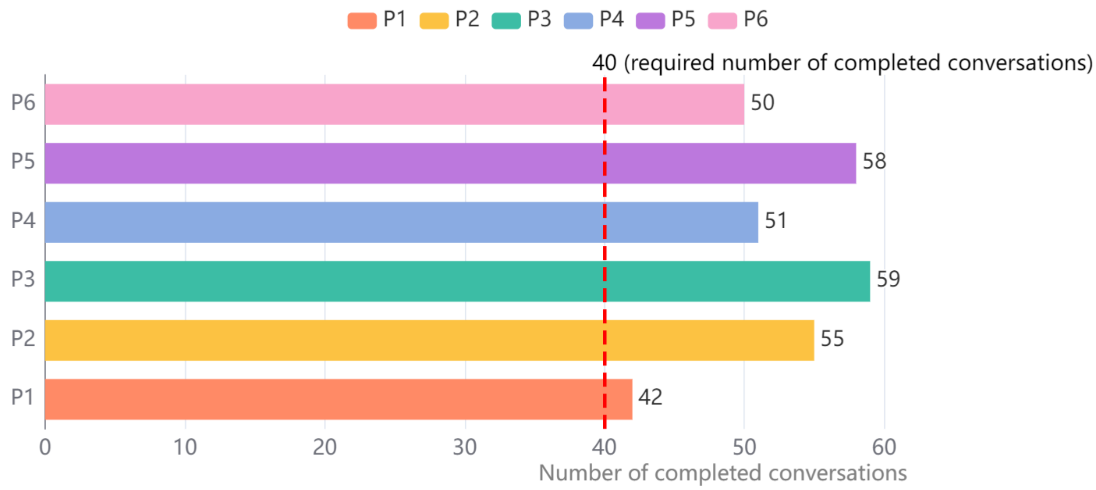


Figure 8: Number of completed conversations by each HFA child.

Audio records of post-test interviews: Audio records of the interviews were transcribed and subsequently subjected to thematic analysis by two researchers. This analysis aimed to identify the perspectives, suggestions, and concerns voiced by parents regarding the utilization of generative AI in assisting children with HFA.

Observation Record sheet during the training period: The observation record sheet completed by parents containing the six children's facial expressions, language, and behavior during the 10-day training is used to evaluate the children's engagement in a thematic analysis. Two autism intervention specialists independently coded the data and reached a consensus through discussion, resulting in five themes relevant to engagement. The inter-coder reliability was verified by the Cohen's Kappa measurement, with scores for each category exceeding 0.76 (maximum: 0.896, minimum: 0.799), indicating that the inter-coder reliability lies between "substantial" and "perfect" [46].

7 FINDINGS

We organized our findings from the user study in three aspects: (1) HFA children exhibited a high level of completion and a positive level of engagement during the interactions with EmoEden, (2) HFA children's abilities of emotion recognition and expression have been improved after using EmoEden continuously for ten days, (3) generative AI exhibits strong comprehension, enhanced personalization and diversified generation in assisting HFA children's emotional learning. Since none of the participants received emotional learning interventions during the user study period, it is unlikely that these improvements in emotion recognition and response were due to other interventions outside the use of EmoEden.

7.1 HFA Children's Completion and Engagement with EmoEden

Overall, the participants exhibited a high level of completion when utilizing EmoEden for emotional training. Although each child was required to complete conversations 4 times per day for 10 days (40 times total), almost all the children completed above 40 required conversations. Each child completed 52.5 conversations on average (min = 42, max = 59), as shown in Figure 8. P1 withdrew from the study on day 7 due to physical illness, nonetheless, we retained his data for subsequent analysis given that P1 had already completed more than 40 conversations.

In addition, the thematic analysis results of the observation record (as shown in Table 9) demonstrated that HFA children exhibited five main behaviors during the 10-day training, with imitation [70], attraction [47], empathy [48], sharing [49], considered as evidence of active engagement. According to the record sheets, almost all HFA children imitated the facial expressions, language, or behavior of the characters sometimes during training. It was also observed that five out of six children exhibited unique responses to certain words in the conversations, three attempted to independently explore EmoEden's specific functions, and three even proactively initiated the training, implying that they were attracted to EmoEden. Four children demonstrated empathy behaviors towards the characters, such as laughing, sighing, or expressing indignation on their behalf. It is worth noting that two children actively shared their conversations with their parents, which may imply that EmoEden brought them an impressive positive experience. Due to network connection failures and unsatisfied responses, two children experienced negative experiences of distraction. Overall, results revealed a positive level of children's engagement with EmoEden.

Children's high level of completion and a positive level of engagement in using EmoEden serve as a foundational basis for subsequent analysis. Additionally, this reflects the usability and attractiveness of EmoEden to HFA children, implying the potential of using EmoEden as a tool for daily emotional training.

Table 9: Theme analysis result of HFA children's behaviors during 10-day training.

Theme	Subtheme	Participant	Example
Imitation	Facial expressions imitation	P1, P2, P3, P5, P6	On Day 2, after seeing the expression of the character change from a smile to a frown, P2's expression also changed from a smile to a frown.
	Language imitation	P1, P3, P4, P5, P6	On Day 3, P5 imitated the character's tone and repeated the sentences involved in conversations spontaneously, such as "understanding and embracing all friendships with heart."
	Action associations	P4, P5	On Day 4, when P4 heard the character picking up dropped potato chips in the supermarket, he made a grasping motion with his hand.
Attraction	Special reactions to certain content	P1, P2, P3, P4, P6	On Day 9, when P3 heard word "coconut" appearing in the conversation, he smiled, clapped hands, and said "delicious".
	Exploration	P1, P3, P5	On Day 5, P1 attempted to click and drag on the circus depicted in the context image to further view.
	Proactive training initiation	P3, P4, P6	On Day 5, P6 ran over for the training actively after hearing the voice of the character.
Empathy	-	P1, P4, P5, P6	On Day 7, when P4 heard the character being cut in line, he angrily tapped the table and kicked his legs.
Sharing	-	P4, P6	On Day 9, P6 initiatively shared with mom that the character ate braised pork.
Distraction	Impatience to latency	P1, P4	On Day 1, P4 ran away while waiting for the long response latency caused by the poor network connection.
	Boredom to similar content	P3, P4	On Day 2, when P3 heard the character narrating an experience similar to a previous one, he complained, "you do something like that again."

7.2 Improvement in Emotional Recognition and Expression Abilities of HFA Children

We conducted a comprehensive evaluation to assess the enhancement of participants' ability in emotional recognition and emotional expression. For emotional recognition, we calculated the accuracy of the participants' recognition of characters' emotions during conversations. For emotional expression, we evaluated the changes in the children's responses to the characters across three dimensions: emotional richness, emotional diversity, and content relevance [50, 51]. Post-test interviews provided additional insights into the emotional behaviors of these children in their daily lives after training, while implying EmoEden's positive impact on their emotional recognition and expression abilities over a period of time.

7.2.1 Improvement in HFA Children's Emotional Recognition. Considering that the pre-test and post-test were both conducted at the institution while the training took place at home, we analyzed the data of the pre-test and post-test and the data during the training separately.

Paired samples t-test on the data of the pre-test and post-test demonstrated a significant improvement ($t\ (5) = -8.883$, $p = 0.0003$) in the accuracy of HFA children's emotion recognition from pre-test ($M = 4.833$, $SD = 1.722$) to post-test ($M = 10.333$, $SD = 1.366$). As shown in Figure 9 (a), P3 and P6 exhibited the most significant improvements by achieving an impressive 58% increase in accuracy. P1, P2, and P5 followed closely, with improvements of 50%, 42%, and 42%, respectively. P4, who already demonstrated a proficient level of emotion recognition in the pre-test, only exhibited a 25% improvement.

An ANOVA was conducted on the data from the 10-day training to assess whether the accuracy of HFA children's emotion recognition increases with the progression of the training. It is worth noting that, although almost all children voluntarily completed extra training sessions, our data collection focused solely on the number of accurate responses within the required four conversations (containing a total of 12 questions) to ensure data consistency. The result confirmed the significant positive impact ($F\ (9,47) = 3.368$, $p=0.002$, Partial $\eta^2=0.411$) of the training duration on the accuracy of emotion recognition, with no violation of the assumptions of normality and homoscedasticity. Figure 9 (b) provides a more intuitive description of the growth trend of the accuracy over the training period.

Overall, the above results suggest that the continuous use of EmoEden improved the emotional recognition abilities of HFA children.

7.2.2 The Impact of EmoEden on HFA Children's Emotional Expression. Emotional expression of HFA children may be affected by personal preferences [52], we used indicators from previous studies: emotional richness [51], emotional diversity [52], and content correlation [50] to evaluate whether their emotional expression ability has improved, rather than whether the content of their expression is appropriate. Text analysis is used to quantitatively evaluate the above three indicators.

Emotional Richness: Emotional richness is evaluated by the total number of all emotion-related words (also called emotion tokens) in the responses of each participant [51]. We first extracted children's responses to characters' emotions from the conversation logs recorded by the system. Then we performed word segmentation by using Stanford CoreNLP library [53] to prepare for subsequent

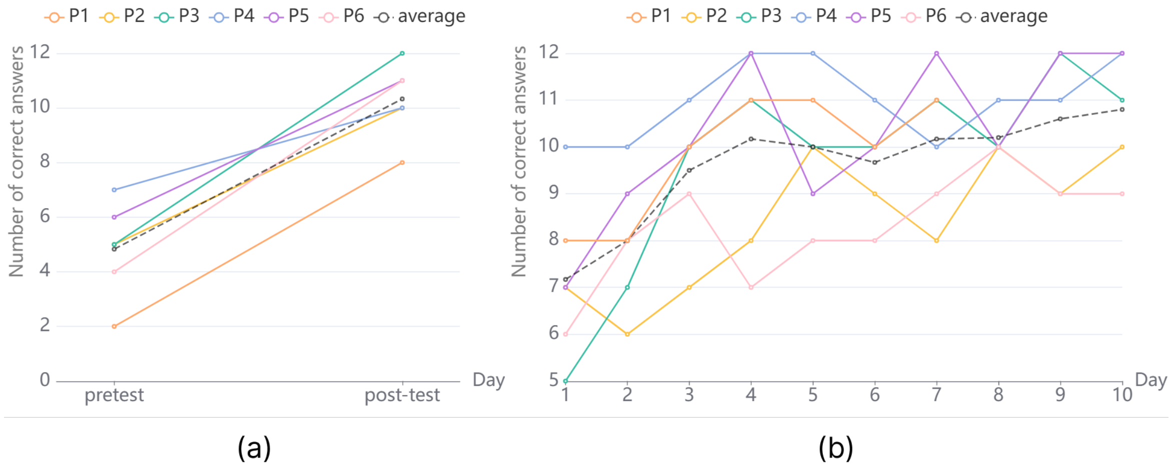


Figure 9: All HFA children exhibited an improvement in the accuracy of emotion recognition. (a) The result of the paired samples t-test shows a significant difference in the accuracy of HFA children's emotion recognition between pre-test and post-test. (b) The growth trend of the accuracy of HFA children's emotion recognition over the training period.

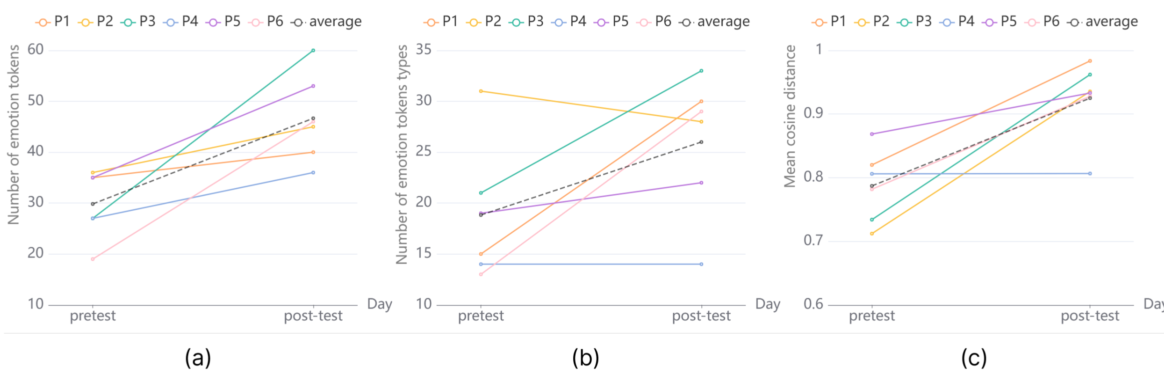


Figure 10: The use of EmoEden had an overall positive impact on the expression of HFA children. (a) The result of the paired samples t-test shows a significant difference in the number of emotion tokens between the pre-test and post-test. (b) The result of the paired samples t-test shows no significant difference in the number of emotion token types between the pre-test and post-test. (c) The result of the paired samples t-test shows a significant difference in the mean cosine distance between pre-test and post-test.

analysis. Considering that the emotional expression in Eastern cultures, especially in Chinese, may be obscure (e.g., you are a talented person) rather than directly involving emotional vocabulary (e.g., I admire you), we performed extraction and tallying based on DUTIR, the largest Chinese sentiment lexicon with 27466 emotional tokens [50, 53, 54]. We calculated the number of emotion tokens in the pre-test and the post-test by each child, and conducted a paired t-test. As shown in Figure 10(a), the results showed a significant difference ($t(5) = -2.954$, $p = 0.032$) in the number of emotion tokens between pre-test ($M = 29.833$, $SD = 6.706$) and post-test ($M = 44.500$, $SD = 8.432$), which indicated an enhancement in emotional richness in the children's expressions.

Emotional Diversity: Emotional diversity is evaluated by the number of different emotion tokens (also called emotion token types) in the responses of each participant [51]. As shown in Figure 10(b), the results of the paired t-test showed there is no significant difference ($t(5) = -2.145$, $p = 0.085$) in the number of emotion token types between pre-test ($M = 18.833$, $SD = 6.706$) and post-test ($M = 26.000$, $SD = 6.899$). This can be explained by the individual differences among HFA children. Some children (P2, P4) may prefer to use the emotion tokens they are already familiar with, so that they did not exhibit diversified emotional expressions. However, other children (P1, P3, and P6) demonstrated more use of newly learned emotion tokens. For instance, P3 learned a new word "generosity" in the conversation of the school scenario on Day 8, and then consciously incorporated it in his expression, stating *"When friends need help, we should support them generously."*

Content Relevance: Content relevance is used to evaluate whether children are focusing on characters' expressions or indulging in self-centered expression. It is calculated as the mean cosine distance between the responses and the corresponding conversation context of each participant. We chose BERT for word embedding [55] on the website (http://wordvec.colorado.edu) [56] to calculate the mean cosine distance of each child in the pre-test and the post-test, and conducted a paired t-test. As shown in Figure 10(c), the results showed a significant improvement ($t(5) = -3.750$, $p = 0.013$) in the mean cosine distance from pre-test ($M = 0.787$, $SD = 0.058$) to post-test ($M = 0.925$, $SD = 0.062$). This may imply that children began to use more vocabulary related to characters' expressions and adjust their expressions according to different contexts. By reviewing the conversation logs, we found that some children even exhibited empathetic expressions. For example, on Day 4, when P3 found the character had difficulties in climbing, he critically pointed out its shortcomings, saying *"You are not suited for climbing. Go do something else."* However, when faced with a similar scenario on Day 8, P3 provided comfort by drawing on his own experiences, saying *"It's okay. You might be afraid of heights like me, but that's normal. You have done well to climb this far."*

In summary, the above results indicate that the use of EmoEden led to the incorporation of more emotion tokens and context-related vocabulary in HFA children's expressions. Due to individual differences, the improvement in the diversity of emotion tokens was reported in a few children but not across all participants.

7.2.3 HFA Children's Emotional Behaviors in their Daily Lives After User Study.

According to post-test interviews, nearly all parents reported enduring improvements in their children's emotional recognition and expression abilities to some extent in their daily lives after the end of the user study.

Emotional Recognition: Four parents mentioned that their children (P1, P4, P5, P6) became more sensitive to the emotions of others. P5's mother noted, *"He started paying attention to his grandmother's emotions and can quickly sense when she's unhappy."* Similarly, P6's mother mentioned, *"After upsetting me, she would covertly observe my expression and attempt to communicate once she sees me calm down, instead of constantly being unruly as before."*

Emotional Expression: Two parents mentioned that their children (P2, P5) became more open to expressing their own emotions rather than keeping them inside. P2's mother observed that her child is trying to verbally express his needs instead of crying and screaming. When faced with an angry mother, P5 would now directly express his feelings, saying, *"Can you please not shout at me? It makes me unhappy"*. Moreover, three parents reported that their children (P3, P5, P6) have begun to mimic the expressions of characters from EmoEden and apply them in real-life situations to express their emotions. P5's mother shared that her child no longer directly snatches their toys when provoked by peers. Instead, he imitated the characters from EmoEden, saying, *"Even though I want your toy, I won't snatch it"* (P5 practiced similar conversation involving disputes with peers while using EmoEden).

7.3 The Performance of Generative AI in Assisting HFA Children in Emotional Learning

By examining the conversation logs, we observed that EmoEden not only effectively achieved its design objectives but also demonstrated some distinctive characteristics of generative AI in assisting HFA children's emotional learning.

7.3.1 Strong Comprehension to HFA Children's Expressions.

We found that EmoEden exhibited a strong comprehension on HFA children's expressions, which may be imperfect or unique. EmoEden can understand incomplete or grammatically incorrect expressions from HFA children and provide appropriate responses. For instance, when P1 noticed the character had a stomachache due to overeating on Day 2, he said, *"no, don't."* EmoEden then comprehended this incomplete input and contextually expanded it to *"Don't eat so much next time, it's not good for your health"*, helping him to express himself effectively. This can be challenging for previous conversation agents, which are reported to generate inappropriate responses when the input lacks sufficient contextual information [57]. However, the capacity of GPT-4 to understand complex language structures and contexts [58] can reduce communication barriers between EmoEden and HFA children, who are often accompanied by language deficits.

In addition, we were surprised to find that EmoEden could even understand the expressions that were entirely created by HFA children themselves. During the communication with P3 on Day 5, EmoEden found P3 liked to use his innovative phrase "gua la (means why)" to express his confusion. Thus, when P3 used this phrase again in subsequent conversations, EmoEden understood it and explained, *"I'm upset because I feel that I am not accepted by their group."* This may be attributed to GPT-4's few-shot in-context learning ability [59]. But for previous designs, they often struggled to generate correlated responses when faced with such inputs involving information that had never appeared before in their training [11, 18]. As mentioned by P4's mother "He likes to talk with it because he believes it understands him." The interviews with parents implied that GPT's strong comprehension may contribute to children's engagement in using EmoEden, which is important for emotional learning [60].

7.3.2 Enhanced personalization.

EmoEden offers personalized updates for the subsequent day's conversational context based on HFA children's daily input, implying its potential to adapt to children's evolving needs. For example, P5's mother input P5's experience of disagreeing with a friend over snack allocation on Day 3. EmoEden then adroitly incorporated the topic of how to resolve conflicts between friends into the conversation in the school scenario on the next day. The therapist highlighted this feature, stating *"incorporating new personalized information may help EmoEden continuously monitor children's long-term changes and provide them with targeted support."*

Such updating of personalized content can be difficult for previous conversational agents, which built dedicated machine learning models based on the data of specific ASD individuals within a specific time period [60, 61]. For those conversation agents that achieve personalization by selecting suitable content from a fixed database according to the user's usage behavior [21, 71], it can be challenging to make the content fully cater to the user's needs. However, these limitations can be addressed with GPT-4's outstanding ability to follow user intentions [14, 62, 63], allowing it to generate content that can be flexibly adjusted through prompts. Considering the individual differences among ASD children, generative AI may improve the effectiveness of emotional learning by providing dynamic and high-quality personalization [60].

7.3.3 Diversified generation. EmoEden was found to be able to extend children's brief input into diversified responses, teaching them to express themselves in various ways. We found that some children (P2, P3, P6) repeatedly clicked the "extend" button in EmoEden after inputting their expressions to obtain various extensions, and chose a suitable one as their final submission. As mentioned by P2's mother *"Sometimes, when he feels like the current extension doesn't capture what he meant, he clicks to replace it."* Such various expression support is due to GPT-4's capacity to generate nuanced and diverse responses [64], which was lacking in previous designs [72].

However, such diverse extensions also resulted in some unsatisfactory responses. Like many applications driven by LLMs [14], EmoEden occasionally introduced unrealistic content to fill in details. One case occurred when P2 shared, "I found an interesting book", EmoEden imaginatively expanded it into *"I found a captivating picture book, which took me into a fantasy world full of adventurous fun."* In fact, it was a popular science book. Nonetheless, only a few parents mentioned the negative impact of this hallucination of LLM in the interview, as P2's mother reported, *"He sometimes gets upset over these inaccuracies."*

8 DISCUSSION

8.1 The Potential Impact of Using Generative AI in Supporting Children with HFA

Despite the potential benefits of integrating generative AI into emotional learning for children with HFA, the ideal approach to this integration remains uncertain [15]. In this section, we discuss the advantages and potential risks of this integration from a macro perspective.

In addition to the abilities of strong comprehension, enhanced personalization and diversified generation, generative AI also improves the accessibility of HFA children's emotional training. As some parents highlighted, generative AI could be used in at-home training to complement human intervention, making it easier for families to integrate the training into their daily routines and providing more equitable training opportunities for those facing difficulties accessing traditional face-to-face therapy. Furthermore, LLMs and text-to-image tools enable users to control the generation of text or visual content using prompts without acquiring specific datasets or building complex deep-learning models. This provides more opportunities for HFA stakeholders to use AI in supporting HFA children. For example, therapists can rapidly generate tailored educational materials, autism experts can effortlessly develop AI-assisted therapeutic tools, and parents or caregivers can actively engage in HFA children's learning using the training data analyzed by AI. Moreover, general-purpose LLMs can effectively handle multiple tasks, making it possible to build integrated tools that combine teaching, assistive expression, and free communication. In the past, these functions typically required separate special education teachers, therapists, augmentative and alternative communication tools, and chatbots to accomplish.

Generative AI may change daily interactions between parents and children in home training where therapists are absent. According to our interviews, since EmoEden requires parents to provide personalized information, parents became more attuned to their children's daily particulars and emotional challenges, which were often overlooked before the study. Furthermore, as mentioned by parents, EmoEden can relieve some of the communication pressure on parents and enhance their understanding of children's needs and emotional states. This implies that generative AI may serve as a mediator in parent-child interactions and help to build a stronger bond within the family.

However, the utilization of generative AI in emotional training for HFA children does carry certain risks. Since it remains unclear whether the hallucination of LLMs would impact negatively on HFA children, some parents expressed their desire for supervision of the conversation content. This study has demonstrated that EmoEden can enhance children's emotional recognition and expression abilities, creating a persisting effect for several days after training. However, considering the complexity of interactions between HFA children and real individuals, it is not sufficient to claim that EmoEden can replace human intervention. To ensure more effective collaboration between AI-assisted emotional training and human interventions in equipping HFA children with the skills to navigate the real world effectively [11], it is crucial to emphasize the balance between AI-based training and face-to-face communications.

8.2 Design recommendations

We sought to summarize the feedback and experience gained from this study and proposed some design recommendations on applying generative AI in supporting HFA children.

8.2.1 Fine-grained personalized settings. Most participants praised EmoEden's ability to generate, extend, and provide feedback based on personal preferences. As P2's mother said, "Whenever the conversation involves his favorite cartoon 'Cute Chicken Team', he claps excitedly." Nevertheless, participants still raised some expectations. Some of them hope that EmoEden has a memory, as they mentioned that although it can understand and remember what they share in a specific conversation, it forgets their interactions from a few days ago. Moreover, we found that EmoEden's narrative style received varying feedback. Some older and more capable children (P3, P5) found it to be somewhat formulaic, while others expected more straightforward language. Therefore, we encourage future work to harness the customization capabilities of generative AI by adding more personalized parameters.

8.2.2 Including HFA children as design partners. Including HFA children as partners in the design process is challenging due to their difficulties in language expression and cognition. In this study, we established design goals by observing HFA children's challenges and preferences and iterated the prototype based on their feedback to ensure that the design aligns with their needs. In this study, we were unable to include children as design partners, therefore, a large portion of the information we gathered came from parents and experts, whose opinions may inherently carry biases. For example, parents have indicated that "I hope he can…" or "he should become…" during interviews. Given the technical affordances of generative AI for designing and developing assistance tools, future studies can involve HFA children in the design process. For instance, text-to-image tools can provide visual materials for HFA children as design inspiration and LLMs can assist them in expressing design suggestions.

8.2.3 Respecting the Autonomy of HFA Children in Emotional Expression. Although HFA children face challenges in expressing emotions, it is essential to strike a balance between preserving their autonomy and offering appropriate guidance when designing conversational companions or communication aids for them. It is worth emphasizing that EmoEden only evaluates whether HFA children accurately recognize other's emotions but does not judge the appropriateness of their emotional expression. EmoEden provides humanoid feedback to help HFA children understand how their expression may affect others and offers various extensions for children to choose or submit their original expressions, helping them express their needs while preserving their individuality as much as possible. Nevertheless, LLMs rely on probability distributions to select the most likely characters as the next output [65], and children's needs may be special. This implies a potential question posed by P6's mother, "will all children's expressions tend to become homogenized by continually using AI-assist system?" Therefore, we strongly recommend preserving adequate autonomy space for HFA children in future designs. We believe that generative AI can serve as a bridge for communication between HFA individuals and the outside world, augment the personalized expression of HFA children, or help them understand others' expressions, rather than replace their expressions.

8.3 Limitations and Future Work

Although we have initially explored generative AI's potential to assist emotional learning for HFA children, there are still some limitations. It should be clarified that in EmoEden, the primary interactive content is generated in real-time by GPT-4, while visual elements pre-generated by Midjourney are non-interactive due to the latency of image generation. Thus, future research could explore text-to-image tools to generate animated character expressions and scenes in real-time to further improve the interaction between children and EmoEden. Additionally, potential biases in the views and expectations of parents and experts, and the lack of voices from HFA children in the formative study may contribute to limitations in the design of EmoEden.

One future direction is the embodiment of EmoEden. To cater to HFA children's preference for tangible objects [66] and provide them with a real-world partner for communication, therapists suggested integrating EmoEden into a physical device, such as toys or smart speakers. Another promising direction is the incorporation of additional user-relevant parameters into EmoEden, enabling it to generate dialogue variations in terms of complexity and vocabulary that better align with the linguistic and cognitive capabilities of HFA individuals.

9 CONCLUSION

In our study, we introduced EmoEden, an AI-driven assistive emotion training tool for children with HFA, which integrates generative AI to enhance emotion recognition and expression skills in HFA children through personalized conversations.

A 22-day user study was conducted with six children to assess the effectiveness of EmoEden. The results showed that HFA children demonstrated a positive level of engagement, exhibited enhancements in their accuracy of emotion recognition, and incorporated diverse emotion tokens and context-related vocabulary in their expressions. And post-training observations revealed that these improvements persisted beyond the training environment. HFA children became more attuned to the emotions of others, displayed a greater willingness to express their own emotions, and actively applied the knowledge acquired from EmoEden conversations in real-life situations.

Additionally, our study shed light on the strengths, potential opportunities, and risks of leveraging Generative AI in supporting emotional learning among HFA children. These insights hold valuable implications for the future development and deployment of generative AI-driven tools in the realm of HFA children's emotional learning.

ACKNOWLEDGMENTS

We extend our warmest thanks to Starvinci for supporting the work. We also thank this project's child and adult collaborators for their enthusiasm and the anonymous reviewers for their contributions. This work is supported by The Ng Teng Fong Charitable Foundation in the form of ZJU-SUTD IDEA Grant (188170-11102).

REFERENCES

[1] Nicholas, J. S., Charles, J. M., Carpenter, L. A., King, L. B., Jenner, W., & Spratt, E. G. (2008). Prevalence and Characteristics of Children With Autism-Spectrum Disorders. Annals of Epidemiology, 18(2), 130–136. https://doi.org/10.1016/j.annepidem.2007.10.013

[2] American Psychiatric Association, A. P., & American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders: DSM-IV (Vol. 4). Washington, DC: American psychiatric association..

[3] Stichter, J. P., O'Connor, K. V., Herzog, M. J., Lierheimer, K., & McGhee, S. D. (2012). Social Competence Intervention for Elementary Students with Aspergers Syndrome and High Functioning Autism. Journal of Autism and Developmental Disorders, 42(3), 354–366. https://doi.org/10.1007/s10803-011-1249-2.

[4] Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A Social Adjustment Enhancement Intervention for High Functioning Autism, Asperger's Syndrome, and Pervasive Developmental Disorder NOS. Journal of Autism and Developmental Disorders, 34(6), 649–668. https://doi.org/10.1007/s10803-004-5286-y.

[5] Rao, P. A., Beidel, D. C., & Murray, M. J. (2008). Social Skills Interventions for Children with Asperger's Syndrome or High-Functioning Autism: A Review and Recommendations. Journal of Autism and Developmental Disorders, 38(2), 353–361. https://doi.org/10.1007/s10803-007-0402-4.

[6] Szatmari, Peter, G. Bartolucci, and Rebecca Bremner. 1989. 'Asperger's Syndrome and Autism: Comparison of Early History and Outcome'. Developmental Medicine & Child Neurology 31(6): 709–20..

[7] Williams White, S., Keonig, K., & Scahill, L. (2007). Social Skills Development in Children with Autism Spectrum Disorders: A Review of the Intervention Research. Journal of Autism and Developmental Disorders, 37(10), 1858–1868. https://doi.org/10.1007/s10803-006-0320-x.

[8] Bauminger, N. (2002). The Facilitation of Social-Emotional Understanding and Social Interaction in High-Functioning Children with Autism: Intervention Outcomes. Journal of Autism and Developmental Disorders, 32(4), 283–298. https://doi.org/10.1023/A:1016378718278.

[9] Kokina, A., & Kern, L. (2010). Social Story™ Interventions for Students with Autism Spectrum Disorders: A Meta-Analysis. Journal of Autism and Developmental Disorders, 40(7), 812–826. https://doi.org/10.1007/s10803-009-0931-0.

[10] Gagan, I. T., Matias, M. A. M., Tan, I., Vinco, C. M., & Ong, E. (2023). Preparing Children with Level 1 ASD for Social Interactions through Storytelling with Amy: An Exploratory Study. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 1–7. https://doi.org/10.1145/3544549.3585916.

[11] Li, X., Zhong, H., Zhang, B., & Zhang, J. (2020). A General Chinese Chatbot Based on Deep Learning and Its' Application for Children with ASD. International Journal of Machine Learning and Computing, 10(4), 519–526. https://doi.org/10.18178/ijmlc.2020.10.4.967.

[12] Shumanov, M., & Johnson, L. (2021). Making conversations with chatbots more personalized. Computers in Human Behavior, 117, 106627. https://doi.org/10.1016/j.chb.2020.106627.

[13] Muller, M., Chilton, L. B., Kantosalo, A., Martin, C. P., & Walsh, G. (2022). GenAICHI: Generative AI and HCI. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, 1–7. https://doi.org/10.1145/3491101.3503719.

[14] Valencia, S., Cave, R., Kallarackal, K., Seaver, K., Terry, M., & Kane, S. K. (2023). “The less I type, the better”: How AI Language Models can Enhance or Impede Communication for AAC Users. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3544548.3581560.

[15] Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. Proceedings of the 22nd Annual ACM Interaction Design and Children Conference, 470–474. https://doi.org/10.1145/3585088.3593867.

[16] Mercer, J. (2017). Examining DIR/Floortime™ as a Treatment for Children With Autism Spectrum Disorders: A Review of Research and Theory. Research on Social Work Practice, 27(5), 625–635. https://doi.org/10.1177/1049731515583062.

[17] Ploog, B. O., Scharf, A., Nelson, D., & Brooks, P. J. (2013). Use of Computer-Assisted Technologies (CAT) to Enhance Social, Communicative, and Language Development in Children with Autism Spectrum Disorders. Journal of Autism and Developmental Disorders, 43(2), 301–322. https://doi.org/10.1007/s10803-012-1571-3.

[18] Catania, F., Spitale, M., & Garzotto, F. (2023). Conversational Agents in Therapeutic Interventions for Neurodevelopmental Disorders: A Survey. ACM Computing Surveys, 55(10), 209:1-209:34. https://doi.org/10.1145/3564269.

[19] Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., Lefeber, D., & David, D. O. (2012). Using the social robot probo as a social story telling agent for children with ASD. Interaction Studies, 13(3), 348–372. https://doi.org/10.1075/is.13.3.02van.

[20] Ali, M. R., Razavi, S. Z., Langevin, R., Al Mamun, A., Kane, B., Rawassizadeh, R., Schubert, L. K., & Hoque, E. (2020). A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons. Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, 1–8. https://doi.org/10.1145/3383652.3423900.

[21] Vijayan, A., Janmasree, S., Keerthana, C., & Baby Syla, L. (2018). A Framework for Intelligent Learning Assistant Platform Based on Cognitive Computing for Children with Autism Spectrum Disorder. 2018 International CET Conference on Control, Communication, and Computing (IC4), 361–365. https://doi.org/10.1109/CETIC4.2018.8530940.

[22] Antona, M., & Stephanidis, C. (Eds.). (2023). Universal Access in Human-Computer Interaction: 17th International Conference, UAHCI 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23–28, 2023, Proceedings, Part I (Vol. 14020). Springer Nature. https://doi.org/10.1007/978-3-031-35897-5.

[23] Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. The International Journal of Management Education, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790.

[24] Alessa, A., & Al-Khalifa, H. (2023). Towards Designing a ChatGPT Conversational Companion for Elderly People (arXiv:2304.09866). arXiv. https://doi.org/10.48550/arXiv.2304.09866.

[25] Montagna, S., Ferretti, S., Klopfenstein, L. C., Florio, A., & Pengo, M. F. (2023). Data Decentralisation of LLM-Based Chatbot Systems in Chronic Disease Self-Management. Proceedings of the 2023 ACM Conference on Information Technology for Social Good, 205–212. https://doi.org/10.1145/3582515.3609536.

[26] Ruskov, M. (2023). Grimm in Wonderland: Prompt Engineering with Midjourney to Illustrate Fairytales (arXiv:2302.08961). arXiv. https://doi.org/10.48550/arXiv.2302.08961.

[27] Chen, T., Lee, C., Mindel, J. R., Elhaouij, N., & Picard, R. (2023). Closer Worlds: Using Generative AI to Facilitate Intimate Conversations. Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3544549.3585651.

[28] Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2016). Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated With Increased Symptom Severity. Biological Psychiatry, 79(8), 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032.

[29] Ludlow, A. K., Wilkins, A. J., & Heaton, P. (2008). Colored overlays enhance visual perceptual performance in children with autism spectrum disorders. Research in Autism Spectrum Disorders, 2(3), 498–515. https://doi.org/10.1016/j.rasd.2007.10.001.

[30] Meadan, H., Ostrosky, M. M., Triplett, B., Michna, A., & Fettig, A. (2011). Using Visual Supports with Young Children with Autism Spectrum Disorder. TEACHING Exceptional Children, 43(6), 28–35. https://doi.org/10.1177/004005991104300603.

[31] Eisen, S., Taggart, J., & Lillard, A. S. (2023). Children Prefer Familiar Fantasy, but not Anthropomorphism, in Their Storybooks. Journal of Cognition and Development, 24(1), 129–141. https://doi.org/10.1080/15248372.2022.2144317.

[32] Andriella, A., Siqueira, H., Fu, D., Magg, S., Barros, P., Wermter, S., Torras, C., & Alenyà, G. (2021). Do I Have a Personality? Endowing Care Robots with Context-Dependent Personality Traits. International Journal of Social Robotics, 13(8), 2081–2102. https://doi.org/10.1007/s12369-020-00690-5.

[33] Elyoseph, Z., Hadar-Shoval, D., Asraf, K., & Lvovsky, M. (2023). ChatGPT outperforms humans in emotional awareness evaluations. Frontiers in Psychology, 14. https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1199058.

[34] Hariri, W. (2023). Unlocking the Potential of ChatGPT: A Comprehensive Exploration of its Applications, Advantages, Limitations, and Future Directions in Natural Language Processing (arXiv:2304.02017). arXiv. https://doi.org/10.48550/arXiv.2304.02017

[35] Whyte, E. M., Smyth, J. M., & Scherf, K. S. (2015). Designing Serious Game Interventions for Individuals with Autism. Journal of Autism and Developmental Disorders, 45(12), 3820–3831. https://doi.org/10.1007/s10803-014-2333-1.

[36] Kim, B., Kim, S.-I., Park, S., Yoo, H. J., Hong, H., & Han, K. (2023). RoutineAid: Externalizing Key Design Elements to Support Daily Routines of Individuals with Autism. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3544548.3581048.

[37] Cha, I., Kim, S.-I., Hong, H., Yoo, H., & Lim, Y. (2021). Exploring the Use of a Voice-based Conversational Agent to Empower Adolescents with Autism Spectrum Disorder. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3411764.3445116.

[38] Williams, J. H. G., Massaro, D. W., Peel, N. J., Bosseler, A., & Suddendorf, T. (2004). Visual–auditory integration during speech imitation in autism. Research in Developmental Disabilities, 25(6), 559–575. https://doi.org/10.1016/j.ridd.2004.01.008.

[39] Santos, K.-A., Ong, E., & Resurreccion, R. (2020). Therapist vibe: Children’s expressions of their emotions through storytelling with a chatbot. Proceedings of the Interaction Design and Children Conference, 483–494. https://doi.org/10.1145/3392063.3394405

[40] Gottman, J. M., Katz, L. F., & Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. Journal of Family Psychology, 10(3), 243–268. https://doi.org/10.1037/0893-3200.10.3.243.

[41] Sansosti, F. J., Powell-Smith, K. A., & Kincaid, D. (2004). A Research Synthesis of Social Story Interventions for Children With Autism Spectrum Disorders. Focus on Autism and Other Developmental Disabilities, 19(4), 194–204. https://doi.org/10.1177/10883576040190040101.

[42] Feng, S., & Chen, C. (2023). Prompting Is All You Need: Automated Android Bug Replay with Large Language Models (arXiv:2306.01987). arXiv. https://doi.org/10.48550/arXiv.2306.01987

[43] Chen, L., Zaharia, M., & Zou, J. (2023). How is ChatGPT’s behavior changing over time? (arXiv:2307.09009). arXiv. https://doi.org/10.48550/arXiv.2307.09009

[44] Deshpande, A., Murahari, V., Rajpurohit, T., Kalyan, A., & Narasimhan, K. (2023). Toxicity in ChatGPT: Analyzing Persona-assigned Language Models (arXiv:2304.05335). arXiv. https://doi.org/10.48550/arXiv.2304.05335.

[45] Gray, C. A., & Garand, J. D. (1993). Social Stories: Improving Responses of Students with Autism with Accurate Social Information. Focus on Autistic Behavior, 8(1), 1–10. https://doi.org/10.1177/108835769300800101

[46] Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159–174. https://doi.org/10.2307/2529310.

[47] Arshad, N. I., Hashim, A. S., Mohd Ariffin, M., Mohd Aszemi, N., Low, H. M., & Norman, A. A. (2020). Robots as Assistive Technology Tools to Enhance Cognitive Abilities and Foster Valuable Learning Experiences among Young Children With Autism Spectrum Disorder. IEEE Access, 8, 116279–116291. https://doi.org/10.1109/ACCESS.2020.3001629.

[48] Witchel, H. J., Santos, C. P., Ackah, J. K., Tee, J., Chockalingam, N., & Westling, C. E. I. (2016). The Complex Relationship Between Empathy, Engagement and Boredom. Proceedings of the European Conference on Cognitive Ergonomics, 1–4. https://doi.org/10.1145/2970930.2970935.

[49] Kim, S. Y., Rispoli, M., Lory, C., Gregori, E., & Brodhead, M. T. (2018). The Effects of a Shared Reading Intervention on Narrative Story Comprehension and Task Engagement of Students with Autism Spectrum Disorder. Journal of Autism and Developmental Disorders, 48(10), 3608–3622. https://doi.org/10.1007/s10803-018-3633-7.

[50] Suh, J., Eigsti, I.-M., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative Performance of Optimal Outcome Children and Adolescents with a History of an Autism Spectrum Disorder (ASD). Journal of Autism and Developmental Disorders, 44(7), 1681–1694. https://doi.org/10.1007/s10803-014-2042-9.

[51] Le Hoang, H. T., & Grégoire, J. (2021). Understanding emotion awareness and emotion vocabulary in early primary school Vietnamese children: Quantitative and qualitative analyses. PsyCh Journal, 10(1), 47–61. https://doi.org/10.1002/pchj.409.

[52] Dewaele, J.-M., & Pavlenko, A. (2002). Emotion Vocabulary in Interlanguage. Language Learning, 52(2), 263–322. https://doi.org/10.1111/0023-8333.00185.

[53] Guan, A., Chen, J., & Tang, T. Y. (2018). ‘Voices of Autism’: Sentiment Analysis in Three Chinese Websites on Nonverbal Autistic Children. Proceedings of the 2018 Artificial Intelligence and Cloud Computing Conference, 115–120. https://doi.org/10.1145/3299819.3299838.

[54] Wang, Y., Zhang, Y., & Liu, B. (2017). Sentiment Lexicon Expansion Based on Neural PU Learning, Double Dictionary Lookup, and Polarity Association. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 553–563. https://doi.org/10.18653/v1/D17-1059.

[55] S. Dennis. (2014) 'How to Use the LSA Web Site', in Handbook of Latent Semantic Analysis, Routledge. https://doi.org/10.4324/9780203936399.ch3.

[56] Losh, M., & Gordon, P. C. (2014). Quantifying Narrative Ability in Autism Spectrum Disorder: A Computational Linguistic Analysis of Narrative Coherence. Journal of Autism and Developmental Disorders, 44(12), 3016–3025. https://doi.org/10.1007/s10803-014-2158-y

[57] Hadri, S. A., & Bouramoul, A. (2023). Towards a deep learning based contextual chat bot for preventing depression in young children with autistic spectrum disorder. Smart Health, 27, 100371. https://doi.org/10.1016/j.smhl.2022.100371.

[58] Panda, S., & Kaur, N. (2023). Exploring the viability of ChatGPT as an alternative to traditional chatbot systems in library and information centers. Library Hi Tech News, 40(3), 22–25. https://doi.org/10.1108/LHTN-02-2023-0032.

[59] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. IEEE/CAA Journal of Automatica Sinica, 10(5), 1122–1136. https://doi.org/10.1109/JAS.2023.123618.

[60] C. Yang, G. G. Bear, and H. May. (2018) Multilevel Associations Between School-Wide Social–Emotional Learning Approach and Student Engagement Across Elementary, Middle, and High Schools, School Psychology Review, vol. 47, no. 1, pp. 45–61, https://doi.org/10.17105/SPR-2017-0003.V47-1.

[61] Francese, R., Guercio, A., Rossano, V., & Bhati, D. (2022). A Multimodal Conversational Interface to Support the creation of customized Social Stories for People with ASD. Proceedings of the 2022 International Conference on Advanced Visual Interfaces, 1–5. https://doi.org/10.1145/3531073.3531118.

[62] Shi, Z., Groechel, T. R., Jain, S., Chima, K., Rudovic, O. (Oggi), & Matarić, M. J. (2022). Toward Personalized Affect-Aware Socially Assistive Robot Tutors for Long-Term Interventions with Children with Autism. ACM Transactions on Human-Robot Interaction, 11(4), 39:1-39:28. https://doi.org/10.1145/3526111.

[63] OpenAI. (2023). GPT-4 Technical Report (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

[64] Liu, Z., Yu, X., Zhang, L., Wu, Z., Cao, C., Dai, H., Zhao, L., Liu, W., Shen, D., Li, Q., Liu, T., Zhu, D., & Li, X. (2023). DeID-GPT: Zero-shot Medical Text De-Identification by GPT-4 (arXiv:2303.11032). arXiv. https://doi.org/10.48550/arXiv.2303.11032.

[65] Zhang, C., Zhang, C., Zheng, S., Qiao, Y., Li, C., Zhang, M., Dam, S. K., Thwal, C. M., Tun, Y. L., Huy, L. L., kim, D., Bae, S.-H., Lee, L.-H., Yang, Y., Shen, H. T., Kweon, I. S., & Hong, C. S. (2023). A Complete Survey on Generative AI (AIGC): Is ChatGPT from GPT-4 to GPT-5 All You Need? (arXiv:2303.11717). arXiv. https://doi.org/10.48550/arXiv.2303.11717.

[66] Wu, Q., Yu, C., Chen, Y., Yao, J., Wu, X., Peng, X., & Han, T. (2020). Squeeze the Ball: Designing an Interactive Playground towards Aiding Social Activities of Children with Low-Function Autism. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 1–14. https://doi.org/10.1145/3313831.3376888.

[67] Im, D. S. (2016). Template to Perpetrate: An Update on Violence in Autism Spectrum Disorder. Harvard Review of Psychiatry, 24(1), 14–35. https://doi.org/10.1097/HRP.0000000000000087.

[68] Vasa, R. A., Hagopian, L., & Kalb, L. G. (2020). Investigating mental health crisis in youth with autism spectrum disorder. Autism Research, 13(1), 112–121. https://doi.org/10.1002/aur.2224.

[69] Anderson, L. K. (2023). Autistic experiences of applied behavior analysis. Autism, 27(3), 737–750. https://doi.org/10.1177/13623613221118216

[70] Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry. Journal of Nonverbal Behavior, 27(3), 145–162. https://doi.org/10.1023/A:1025389814290

[71] El Rhatassi, F. E., El Ghali, B., & Daoudi, N. (2023). A Chatbot's Architecture for Customized Services for Developmental and Mental Health Disorders: Autism. In S. Motahhir & B. Bossoufi (Eds.), Digital Technologies and Applications (pp. 134–141). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-29860-8_14

[72] Allouch, M., Azaria, A., & Azoulay, R. (2021). Conversational Agents: Goals, Technologies, Vision and Challenges. Sensors, 21(24), Article 24. https://doi.org/10.3390/s21248448.