RESEARCH ARTICLE

Glitter: Exploring an LLM Virtual Agent for Supporting Practitioners in Behavioral Interventions of Autistic Children

Xin Tong[a,b]†, Liwen He[b,f]†, Zhaowen Deng[d], Weibo Li[e], Ziheng Tang[b,g], Yixuan Li[a,h], Yutong Ren[d,i], Matthew Louis Mauriello[j] and Ming Li[c]*

[a]Hong Kong University of Science and Technology (Guangzhou), Thrust of Computational Media and Arts, No. 1 Duxue Rd., Guangzhou, 511453, China;
[b]Duke Kunshan University, Data Science Research Center , No. 8 Duke Avenue, Kunshan, 215306, China;
[c]Duke Kunshan University, Digital Innovation Research Center, No. 8 Duke Avenue, Kunshan, 215306, China;
[d]Duke Kunshan University, Undergraduate, No. 8 Duke Avenue, Kunshan, 215306, China;
[e]Duke Kunshan University, Electronical and Computer Engineering, No. 8 Duke Avenue, Kunshan, 215306, China;
[f]Tsinghua University, The Future Laboratory, No. 30 Shuangqing Rd., Haidian District, Beijing, 100084, China;
[g]Tsinghua University, School of Mechanical Engineering, No. 30 Shuangqing Rd., Haidian District, Beijing, 100084, China;
[h]Tsinghua University, Shenzhen International Gradute School, Tsinghua Campus, the University Town, Shenzhen, 518055, China;
[i]University of Michigan, School of Information, 500 S State St, Ann Arbor, MI 48109, USA;
[j]University of Delaware, Department of Computer and Information Sciences, 210 S College Ave, Newark, DE 19716, USA

ARTICLE HISTORY
Compiled April 9, 2026

ABSTRACT
Autistic children often exhibit high rates of behavioral issues. While Applied Behavior Analysis (ABA) is a widely recognized framework for behavioral interventions, practitioners face challenges in personalizing and integrating it into diverse real-world contexts. Given the potential of Large Language Models (LLMs) to enhance information access through personalized conversations, we developed *Glitter*, an LLM-powered virtual agent, and investigated its roles, opportunities, and challenges in supporting ABA-integrated interventions. Using a three-phase iterative design process, we identified the needs of practitioners and implemented a system, Glitter, combining a self-directed learning module and an interactive LLM agent. In a one-week field study involving 11 practitioners, we examined system usability, real-world use cases, and practitioner perceptions. Results revealed three primary use cases for Glitter: real-time problem-solving (Inquire & Resolve), strategy refinement (Compare & Improve), and lesson preparation (Prepare & Conclude). Practitioners employed three distinct prompting strategies—single inquiries, rephrased queries, and progressive prompts—to optimize interactions with the agent. This research demonstrates the potential of an LLM agent in providing adaptive, therapist-facing support for behavioral interventions. It offers a novel approach to integrating ABA

† Xin Tong and Liwen He contributed equally to this work and share first authorship.
CONTACT (Submitting Author) Liwen He. Email: buaaheliwen@163.com
CONTACT (Corresponding Author) Ming Li. Email: ming.li369@dukekunshan.edu.cn

principles into digital tools, provides field-based observations on the opportunities and challenges of LLMs in behavioral interventions, and presents design implications for future virtual agents supporting autistic interventions.

KEYWORDS
Autism; Applied Behavior Analysis (ABA); Large Language Models (LLMs); Human–Computer Interaction (HCI); Behavioral interventions

1. Introduction

Autism Spectrum Disorder (ASD) has been commonly introduced as a highly heterogeneous neurodevelopmental disorder through the lens of the medical model of disability (American Psychiatric Association, 2013). More recently, ASD has been redefined by researchers as a "deficit-as-difference" and self-identified by individuals as a type of neurodiversity (Kapp, Gillespie-Lynch, Sherman, & Hutman, 2013). Autistic children[1] usually demonstrate a wide spectrum of varying behaviors and developmental skills in social interaction and communication (American Psychiatric Association, 2013). Furthermore, autistic children frequently experience emotional and behavioral difficulties—commonly described as "problem behaviors" (Dominick, Davis, Lainhart, Tager-Flusberg, & Folstein, 2007) or "challenging behaviors" (J. Bradshaw, Wolfe, Hock, & Scopano, 2022), such as aggression, oppositional actions, or tantrums—across daily life and social interactions (Dominick et al., 2007; J. Matson, 2009; Zaidman-Zait et al., 2014). These behaviors are often difficult to address and can make daily living challenging for autistic children and practitioners, caregivers, or family members (Fox, Vaughn, Wyatte, & Dunlap, 2002).

Early behavioral interventions are essential for improving the well-being of autistic children (Eldevik et al., 2009a; Reichow, Barton, Boyd, & Hume, 2014), addressing behavioral difficulties (Reichow et al., 2014), and reducing practitioners' stress (Tarver et al., 2019). Training and support for those teaching autistic students have therefore become increasingly important (National Research Council, 2001). Although behavioral strategies such as Applied Behavior Analysis (ABA) are considered standard care (Buchanan & Weiss, 2006), both caregivers and practitioners face barriers to accessing, adopting, and integrating these approaches in daily practice (Babalola, Sanguedolce, Dipper, & Botting, 2024). High-quality training equips practitioners with current knowledge and skills for personalized interventions (Curran, Roberts, Gannoni, & Jeyaseelan, 2024), strengthens their confidence (P. Bradshaw, Pickett, van Driel, Brooker, & Urbanowicz, 2021), and improves job satisfaction and professional identity while reducing burnout, supporting the long-term sustainability of educational services (Clarke & Fung, 2022).

Despite these benefits, current non-digital and digital learning and training resources often fall short in supporting practitioners. Common training formats, such as workshops or teacher preparation programs, offer limited flexibility (Morrier, Hess, & Heflin, 2011) and fail to meet the needs of general practitioners with limited expertise in ASD (Chown et al., 2023; Morrier et al., 2011). These formats typically provide one-time, generalized instruction that is difficult to translate into ongoing, context-specific practice, especially when practitioners face rapidly evolving or unpredictable behavioral situations. Additionally, prior research has investigated how technologies

[1]In this paper, we used "autistic children" versus "children with autism" because the autism community preferred being described in identity-first language over person-first language according to prior research (Kenny et al., 2016).

can assist practitioners in learning and delivering interventions (Hamad, Serna, Morrison, & Fleming, 2010; Sam, Cox, Savage, Waters, & Odom, 2020). However, most of these systems are limited to duplicating content and targeting autistic children as end users, with the primary goal of improving their skills (e.g., social (I.-J. Lee & Hsu, 2023), language and speech (Sandbank et al., 2020), and emotional regulation (Tang et al., 2024)), without considering the roles and demands of practitioners. Such tools rarely support practitioners in real time, lack mechanisms for tailoring strategies to unique classroom or clinical contexts, and do not integrate with practitioners' workflows of observation, reflection, and decision-making. Meanwhile, the heterogeneity of ASD (National Research Council, 2001) further complicates training processes, as strategies effective for some children may not work for others (Dominick et al., 2007).

Practitioners consistently prefer personalized, practice-oriented, one-on-one training that provides step-by-step, child-specific guidance (Alexander, Ayres, & Smith, 2015). Although prior work investigated how technologies could assist practitioners in learning and delivering interventions (Hamad et al., 2010; Sam et al., 2020), most are limited to duplication of digital materials in an online website format (Villamin & Luppicini, 2024) and fail to provide dynamic, adaptive support (Hamad et al., 2010). Taken together, existing traditional and digital resources do not adequately address practitioners' needs for flexible, context-sensitive, and personalized technology support, highlighting opportunities to design tools that explicitly position practitioners as primary users.

Recently, the emergence of Large Language Models (LLMs) has shown promise in improving communication and individual access, informing its potential for adoption in general behavioral health (E. C. Stade et al., 2024) and mental wellbeing (Ma, Mei, & Su, 2023). However, there is limited work investigating how LLM agents could be applied to autistic stakeholders (Jang, Moharana, Carrington, & Begel, 2024; McFayden, Bristol, Putnam, & Harrop, 2024; Ren et al., 2024a). While LLM-based systems have been explored in adjacent healthcare and educational contexts, less is known about how practitioner-facing LLM agents can be structured around ABA-informed workflows and used in real-world intervention settings. Therefore, in this work, we aim to further explore technological approaches in supporting stakeholders of autistic children and investigate the use cases, opportunities, and challenges associated with an LLM-based agent for supporting practitioners in ABA-integrated interventions. Rather than evaluating intervention efficacy, this study focuses on how practitioners use and perceive an LLM-based consultation agent in real-world contexts, and what opportunities and limitations emerge from such deployment. Specifically, we aim to answer the following research questions (RQs):

- RQ1: What are the practical use cases and perceived usefulness of an LLM-based agent in supporting practitioners' ABA-integrated interventions for autistic children?
- RQ2: How do practitioners perceive and respond to the visual appearance, speech, and affective responses of an LLM-based agent during ABA-related consultations?
- RQ3: What are the opportunities and challenges of using an LLM-based agent to assist practitioners in ABA-integrated interventions for autistic children?

To answer these research questions, we first adopted a three-round iterative design-research process, involving caregivers and practitioners working with autistic children, to collect topics and understand practitioners' current situations, problems, and expectations regarding technologies. According to the findings, we then developed Glitter, a prototype that provided a self-directed learning module and an LLM agent for assisting practitioners in their behavioral interventions. Next, we conducted a one-week field study with 11 practitioners to further answer these research questions by collecting evidence from practical scenarios. Findings suggested that practitioners provided positive feedback and considered Glitter to be overall trustworthy and useful as a dedicated ASD application. Their major use case scenarios and motivations of the LLM-agent were: **Inquire & Resolve**, **Compare & Improve**, and **Prepare & Conclude**. Practitioners also shared their perspectives about the opportunities and concerns of applying Glitter in practical scenarios. For example, they found crafting effective prompts for the LLM agent to be challenging and expressed concerns about its language expression, including terminology and stylistic nuances. On a positive note, they appreciated Glitter's privacy and security settings. Findings also revealed that practitioners ranked the agent's speech features most highly and considered the intervention content to be critical for information acquisition. Although the avatar's appearance offered practitioners a sense of presence and realness, results showed that none of the visual or affective features influenced practitioners' sense of information acquisition. To conclude, this work contributes:

- a novel approach to integrating the ABA behavioral intervention framework in digital interventions via an iterative design-research process;
- an exploratory understanding of practitioners' prompting strategies and use case scenarios in practical intervention, and their perceptions of potential opportunities and challenges from a field study;
- observational insights into how practitioners perceived the LLM agent's visual appearances, speech, and affective responses during real-world information expression and acquisition; and
- design considerations for future practitioner-facing LLM agents intended to support behavioral intervention practice.

2. Related Research

This section reviews prior work relevant to our study, including behavioral intervention strategies for autistic children, digital tools that support practitioners' learning and daily practice, and recent applications of large language models in psychotherapy and autism-related contexts. Together, these strands provide the background for understanding current practices and technological support in this domain.

2.1. Behavioral Intervention Strategies for Autistic Children

Autistic children often face significant behavioral challenges, including aggression and noncompliance, which can hinder their developmental progress and overall quality of life (L.-C. Lee, Harrington, Louie, & Newschaffer, 2008; J. L. Matson & Nebel-Schwalm, 2007). These challenges frequently contribute to heightened stress and emotional burdens for parents and caregivers (Bonis, 2016). To address these issues, behavioral interventions grounded in *Applied Behavior Analysis (ABA)* have emerged as a widely adopted approach, leveraging operant conditioning principles to promote functional independence (Skinner, 1971).

ABA-based interventions are recognized for their adaptability, enabling personalized, one-on-one implementation to support diverse behavioral needs. Core methodologies within the ABA framework include *Early Intensive Behavioral Interventions (EIBI)* (Vismara & Rogers, 2010), *Verbal Behavior interventions (ABA-VB)* (J. E. Carr & Firth, 2005), *Discrete Trial Training (DTT)* (Smith, 2001), the *Picture Exchange Communication System (PECS)* (Charlop-Christy, Carpenter, Le, LeBlanc, & Kellet, 2002), and *Positive Behavioral Supports (PBS)* (E. G. Carr et al., 2002). Collectively, these approaches are commonly referred to as *ABA therapy* and represent a cornerstone of clinical practice for individuals with ASD. Despite their widespread usage, ABA-based interventions continue to face criticism, particularly regarding their emphasis on observable behavior change and their limitations in supporting more nuanced social-emotional competencies (Lim, 2020; Wilkenfeld & McCarthy, 2020). Nonetheless, ABA remains supported by a robust theoretical and empirical foundation, with meta-analytic evidence demonstrating consistent benefits across cognitive, adaptive, and language domains (Eldevik et al., 2009b; Makrygianni, Gena, Katoudi, & Galanis, 2018; Makrygianni & Reed, 2010; Peters-Scheffer, Didden, Korzilius, & Sturmey, 2011; Reichow, Barton, Boyd, & Hume, 2012; Reichow & Wolery, 2009; Spreckley & Boyd, 2009; Virués-Ortega, 2010).

However, a distinct set of challenges arises when translating these well-established methods into everyday practice. Across service contexts, the supply of trained ABA professionals continues to lag behind demand, and frontline practitioners often report insufficient preparation for delivering high-quality interventions (Mahon, Dunworth, McSharry, Holloway, & Lydon, 2025). In real-world settings, practitioners shoulder substantial cognitive and emotional workload due to the need for continuous case-specific decision-making, moment-to-moment procedural adjustments, and real-time synthesis of behavioral observations (Caldwell, Schreck, Spisak, & Katz, 2025). Maintaining treatment fidelity further amplifies these demands, especially when intervention protocols must be adapted to accommodate each child's unique behavioral profile (Pettersson et al., 2025; Rivard et al., 2025). This inherent tension between fidelity and personalization necessitates ongoing reinterpretation and situational adjustment of standardized techniques (Bromley, 2023).

Building on the well-established theoretical and empirical basis of ABA (Eckes, Buhlmann, Holling, & Möllmann, 2023), as well as feedback gathered from practitioners during pilot studies (see more details in Section 3.1), this work adopts the ABA framework as the guiding principle for the *Glitter* app. Specifically, *Glitter* employs an expanded *Antecedent-Behavior-Consequence (ABC)* model (Lovaas, Schreibman, & Koegel, 1974) and incorporates *DTT* as its primary methodology. This is complemented by additional techniques, including reinforcement, extinction, prompting, video modeling, and PECS (Charlop-Christy et al., 2002; Granpeesheh, Tarbox, & Dixon, 2009; Sandbank et al., 2020; Stahmer, Collings, & Palinkas, 2005). These strategies are designed to help practitioners deliver structured and repetitive teaching sessions within distraction-free, one-on-one environments.

2.2. Digital Tools and Technologies for Practitioners' Self-Directed Learning in Autistic Children's Behavioral Interventions

Recent years have seen an expansion of digital tools developed to support practitioners in ABA-based interventions. Rather than functioning as end-to-end intervention systems, these tools typically strengthen specific operational components of the workflow, such as material preparation, child-facing activity structuring, data management, remote delivery, or self-directed learning, while leaving practitioners responsible for synthesizing information and making procedural decisions (see Table 1).

2.2.1. Tools for preparing instructional materials

These systems assist practitioners in producing digital teaching materials such as DTT stimuli, visual prompts, or activity components (Blair & Shawler, 2020; Mattson et al., 2020). They help reduce preparation time but generally operate independently of the child's behavioral profile or instructional context, and therefore provide limited support for selecting procedures or planning multi-step intervention sequences.

2.2.2. Tools for structuring child-facing task environments

Digital activity schedules, visual supports, and token systems provide predictable routines that help children complete tasks and transitions (Hammond, Morris, Gabrielsen, Smith, & Medsker, 2025; Lindgren, 2025). In most cases, practitioners configure or update systems manually, and the tools do not adapt to moment-to-moment behavioral changes or offer suggestions to modify prompting or reinforcement procedures.

Table 1.: Functional Categories of Digital Tools for Supporting Practitioners.

Category	Primary Function & Key Limitations
Instructional Material Preparation	**Function:** Supports creation of digital DTT stimuli, visual prompts, and teaching materials (Blair & Shawler, 2020; Mattson et al., 2020). **Limitation:** Static content generation; operates independently of the child's real-time behavioral context.
Child-Facing Structured Environments	**Function:** Provides digital activity schedules, visual supports, or token systems (Hammond et al., 2025; Lindgren, 2025). **Limitation:** Requires manual configuration; does not adapt to moment-to-moment behavioral changes.
Data Collection & Monitoring	**Function:** Enables real-time data logging, automated graphing, and progress review (Bak et al., 2021; Slanzi & Fernand, 2024). **Limitation:** Descriptive only; leaves analytic interpretation and procedural adjustments entirely to the practitioner.
Telehealth & Self-Directed	**Function:** Facilitates remote modeling, coaching, and self-paced learning (Sutherland, Trembath, & Roberts, 2018; Wood et al., 2024). **Limitation:** Feedback is often asynchronous or curriculum-based, lacking "just-in-time" decision support for active interventions.
Generic LLMs (e.g., ChatGPT)	**Function:** Provides general Q&A and text generation (McFayden et al., 2024).

Limitation: Lacks domain-specific ABA guardrails (e.g., BABC logic); high risk of hallucination in clinical advice.

2.2.3. Tools for remote guidance and self-directed learning

Telehealth platforms support real-time modeling, feedback, and consultation in home settings (Pomales-Ramos, Tokish, Howard, Straiton, & Ingersoll, 2023; Sutherland et al., 2018). Web-based or mobile learning programs similarly provide practitioners and caregivers with self-paced instructional materials (Glenn, Taiwo, Arbuckle, Riehl, & McIntyre, 2023; Ibañez et al., 2018), and have been shown to increase accessibility and improve the delivery of evidence-based practices (Wood et al., 2024). However, both telecoaching and self-directed modules emphasize content access and skill delivery rather than helping users translate broad principles into individualized decisions across diverse situations (McIntyre & Phaneuf, 2008; Pickard, Rowless, & Ingersoll, 2019).

Across these categories, existing tools support multiple aspects of ABA practice but remain loosely connected and do not form an integrated system for coordinating decisions across situations. This fragmentation suggests an opportunity to explore technologies that better link instructional routines, behavioral information, and procedural choices in a more coherent manner.

2.3. Large-Language Models in Psychotherapies and Autistic Interventions

Researchers have developed various AI-based chatbots, conversational agents, and virtual agents (referred to as 'agent' hereafter) that provide diagnostic assessment information (Ren et al., 2024b), evidence-based evaluation and intervention content (Oh, Lee, Ko, & Choi, 2017), and offer feedback to clinical practitioners (McGreevey III, Hanson III, & Koppel, 2020). With the advancement of LLMs over the past few years, LLM agents can generate and offer mental wellness and behavioral intervention support to users by engaging in open dialogues through prompting, which enables emotionally responsive interactions with users. These agents have promise in assisting with emotional wellbeing (Ferrara, 2022), counselling (Qiu & Lan, 2024), behavioral health (E. C. Stade et al., 2024), and other psychotherapies (E. Stade et al., 2023), most of which showed positive intervention outcomes and experiences in user research (Ma et al., 2023). Nonetheless, LLM-driven healthcare agents are still in their early developmental stages; studies have revealed that users encounter challenges and concerns in adopting these agents, such as generating inappropriate content, inconsistent effectiveness, hallucinations, and context loss across interactions (H. Li, Zhang, Lee, Kraut, & Mohr, 2023).

Stade et al. (E. C. Stade et al., 2024) categorized all possible LLM agents in psychotherapy applications into four major types according to the targeted users: patient-facing, therapist-facing, trainee-facing, and supervisor/consultant-facing. As suggested by E. C. Stade et al. (2024), therapist-facing agents should offer intervention options for the therapist to select. Trainee-centered agents, on the other hand, offer feedback on the trainee's performance qualities, whereas a supervisor/consultant-facing agent could summarize therapy sessions. Most current generative AI agents for autism in the HCI domain are mainly "patient-facing," which provide interventions or educational content directly to the autistic adults or children to support their communication skills (Deng et al., 2024; Jang et al., 2024; Z. Li, Babar, Barry, & Peiris, 2024; Mishra & Welch, 2024) and emotional regulation (Tang et al., 2024), provide emotional support (Jang et al., 2024), generate social stories for learning social skills (Alkinj, Pereira, & Santos, 2022; Vanderborght et al., 2012), and so on. For instance, Jang et al. (2024) explored how autistic adults envisioned using LLM agents to support workplace social communication through an interview study. They found that participants strongly preferred LLM interactions over those with a human confederate, though concerns were raised about the quality of the advice provided by the LLM.

However, few therapist-facing agents in autism interventions have been investigated in practical studies. Existing LLM agents are largely patient- or child-facing or serve as general information portals for parents (McFayden et al., 2024). While general-purpose LLMs (like ChatGPT) can generate broad advice, they are functionally inadequate for professional ABA practice for two key reasons. First, they lack domain-specific analytic logic; they do not inherently follow the structured BABC (Behavior-Antecedent-Behavior-Consequence) framework required for rigorous intervention, often resulting in generic suggestions rather than precise procedural adjustments. Second, they pose a risk of hallucination: without grounded retrieval databases, generic models may generate plausible but clinically inaccurate strategies. Consequently, current agents do not directly address practitioners' need for a system that combines the reasoning capabilities of LLMs with the structured guardrails of ABA protocols to support real-time decision-making. Moreover, few studies have examined how therapist-facing LLM agents can be embedded into practitioners' day-to-day intervention workflows and observed in real-world use.

3. System Design

3.1. System Design Process

This section outlines the system design process, which included three rounds of pilot studies conducted with practitioners (see Figure 1). The objectives of these studies were twofold: (1) to establish a preliminary database identifying relevant topics and areas of focus for the system, and (2) to investigate practitioners' self-learning practices, challenges, and expectations for technological tools to support their work in guiding caregivers.

Insights gained from these studies informed a series of co-design sessions with professional practitioners. These sessions focused on integrating key findings into the system's design features. Following this, a formative evaluation was conducted with practitioners to refine the system through iterative prototype development.

3.1.1. Pilot Study 1: Exploring Practitioner Challenges and Expectations for Supporting Caregivers in Behavioral Interventions

Study Goals and Methods. The first pilot study involved semi-structured interviews with seven caregivers (five females, two males, $M = 35.86$ years, $SD = 2.79$) of autistic children to explore their expectations about learning from professional practitioners (i.e., special educators). Additionally, interviews were conducted with 16 practitioners (13 females, three males, $M = 27.63$ years, $SD = 3.79$) to gather comparable data, facilitating a comparison between the perspectives of caregivers and practitioners. This approach aimed to identify the challenges practitioners encounter when guiding caregivers. The study also explored practitioners' self-learning behaviors, the barriers they

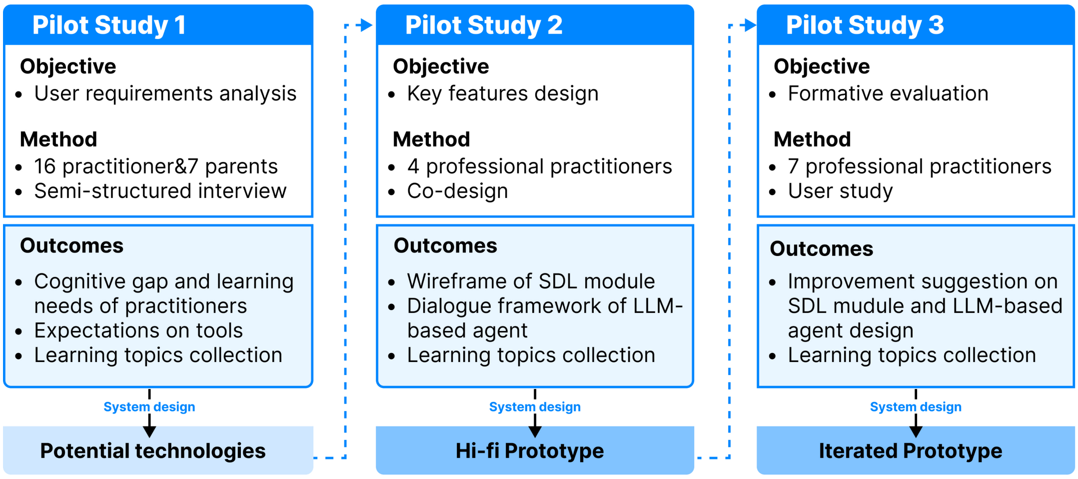

Figure 1. Design process of Glitter system.

face, and their expectations for tools supporting the acquisition and delivery of intervention strategies. Two researchers conducted a thematic analysis of the interview transcripts. To derive the design considerations, we applied an inductive thematic analysis approach (Fereday & Muir-Cochrane, 2006). Two researchers independently coded the interview transcripts, compared codes to resolve discrepancies, and collaboratively grouped them into higher-level themes representing practitioners' learning needs and caregiving challenges. The themes were then iteratively reviewed and mapped to actionable design implications, ensuring that each consideration could be directly traced back to recurring patterns in the qualitative data.

Results. The caregiver interviews revealed that while caregivers often rely on professional practitioners for guidance, many also seek information from online platforms such as WeChat[2], Zhihu[3], or Douyin[4], due to the few opportunities to consult with practitioners. However, these resources are often fragmented, lack systematic explanations, and require significant effort to verify. Caregivers expressed a strong preference for structured, actionable intervention strategies curated by professionals. Although some caregivers consult practitioners before applying new methods, practitioners reported challenges in providing detailed, customized plans due to time constraints, especially in one-to-many service scenarios.

Practitioners faced challenges during their learning processes. They often require access to professional resources, such as textbooks and research articles, to analyze complex behavioral problems. While digital tools, such as ASD-focused Weixin official accounts[5] or expert videos on Douyin, offer quick access to general knowledge, they often fail to provide specific, customizable, and efficient solutions. Furthermore, practitioners' learning needs vary according to their experience. Novices require systematic frameworks that bridge theoretical understanding and practical application, while experienced practitioners seek innovative strategies to refine their expertise. The findings also highlighted a cognitive gap between caregivers and practitioners. While practitioners typically focus on providing actionable strategies, some caregivers noted that understanding theoretical foundations enhances their ability to implement these

[2]WeChat: `https://www.wechat.com`
[3]Zhihu: `https://www.zhihu.com`
[4]Douyin: `https://www.douyin.com`
[5]Weixin official accounts: `https://mp.weixin.qq.com`

strategies effectively.

Based on our findings, the following design considerations are proposed to develop a system that supports practitioners:

- **Wide Range of In-home Intervention Topics**: The system should encompass different in-home intervention topics, including but not limited to social development, communication deficits, and emotional management.
- **Knowledge Tailored to Experience Levels**: The system should provide knowledge resources of varying depths to accommodate practitioners with diverse levels of expertise, enabling both foundational learning and advanced exploration.
- **Theory-to-Practice Knowledge Transformation**: A systematic framework should support the transition from theoretical knowledge to actionable steps, helping practitioners transfer knowledge to caregivers with consistency and clarity.
- **Structured Presentation of Intervention Steps**: The intervention steps should be presented in a clearly organized and replicable format.
- **Personalized and Real-Time Adaptation**: The system should support real-time customization, enabling practitioners to respond promptly to classroom scenarios or caregivers' specific queries.

3.1.2. Pilot Study 2: Co-Designing Intervention Technologies with Practitioners of Autistic Children

After developing an understanding of how parents obtain behavioral intervention information in Pilot Study 1, we then further explored potential technologies and approaches for delivering interventions.

Study Goals and Method. We conducted a co-design session with four professional practitioners (two behavioral analysts/therapists and two special education teachers, three female, one male, $M = 37.50$ years, $SD = 3.32$) to explore how technology could better support the challenges practitioners encounter during behavioral interventions. The session followed a multi-stage participatory design process. First, practitioners were invited to describe their real-world learning and intervention challenges, as well as the types of support they would like digital tools to provide. This exploratory discussion helped surface pain points and expectations without introducing predefined solutions. Second, based on these articulated needs, we introduced a set of preliminary design elements (e.g., possible interaction components and ABA-grounded functional units) to prompt collaborative ideation. Practitioners evaluated these elements, noted missing components, and reflected on how each might fit into their intervention workflows. Third, practitioners and researchers jointly assembled these design elements into an initial interaction-flow framework that represented the sequence of decisions, information access, and guidance steps needed during practice. This collaborative construction clarified how a system might organize prompts, provide feedback, and support reasoning. Finally, the collaboratively generated interaction flow was translated into low-fidelity wireframes. Practitioners reviewed these wireframes and annotated the intended function of each component, offering refinements and highlighting places where emotional assurance, contextual guidance, or personalized suggestions from an LLM-based agent would be particularly valuable. Their feedback formed the co-design artifacts that informed the subsequent development of the prototype.

Results. All participants underscored that the ABA framework is one of the most structural and practical theories they learned and adopted to guide their daily intervention practices. They believed that proficiency in the ABA theory laid the cornerstone for scientifically rigorous interventions. Therefore, participants recommended implementing an ABA-grounded intervention model into the digital application. Practitioners also suggested adopting a self-paced and self-directed learning online platform for novice practitioners, showing them how to implement the progressive steps of Background-Antecedent-Behavior-Consequence (BABC) in diverse intervention topics and real-life cases that practitioners and caregivers commonly find challenging. Based on these discussions, we further sketched out a low-fidelity wireframe and user flow for a mobile self-directed learning module together.

Further, practitioners also addressed the importance of enhancing the overall "learnability, accessibility, flexibility, and approachability of intervention knowledge" and personalizing the instructional guidance and intervention content for each case due to autistic children's diverse abilities, motivational triggers of behavioral issues, and the specific intervention contexts. During the design process, practitioners observed that the agent's responses were often too general. This issue arose because the LLM-based agent's responses heavily rely on the specificity of user input, and users tend to struggle with efficiently organizing the necessary key information when providing prompts. To address this challenge, we plan to implement a structured dialogue framework for the conversation agent. This framework will guide practitioners in formulating precise questions and providing relevant context and character details, ensuring that the agent can generate more personalized and contextually relevant responses. To inform the design of this framework, we collected conversation examples from multiple sources, including interactions between practitioners and caregivers of autistic children, examples of how people typically seek help when encountering intervention challenges, and relevant learning resources. By analyzing these conversations, we identified the key pieces of information that need to be included in the prompts, ensuring that the structured dialogue framework supports practitioners in obtaining relevant and personalized answers from the agent, as well as how to offer behavioral intervention guidance and knowledge in natural conversations. In this study, we also collected the behavioral issues that practitioners frequently encounter and their intervention strategies, following the ABA framework and BABC analysis steps.

3.1.3. Pilot Study 3: Evaluating the Self-Learning Module and the Virtual Agent Design Features with Autistic Children Practitioners

After summarizing the key intervention features and low-fidelity designs from our co-design findings in Pilot Study 2, we implemented the interactive high-fidelity prototype, Glitter, and evaluated it with practitioners in Pilot Study 3 to iterate on and improve the design features and content. In this study, we further collected intervention topics and practical processes, as well as querying scenarios, to supplement our repository for iterating the content and design of the self-directed learning (SDL) module, as well as the intervention strategies of LLM-based agents.

Study Goals and Method. We recruited 7 practitioners from special educational institutions (six female, one male, M = 26.29 years, SD = 4.99) to evaluate the prototype through an in-lab study. We invited all participants to test Glitter for 15-20 minutes and interviewed them afterward to collect their feedback on the general system usability, thoughts on the content and learning flow of the SDL module, and experiences querying the agent.

Results. Most participants thought Glitter was easy to learn and use. Participants

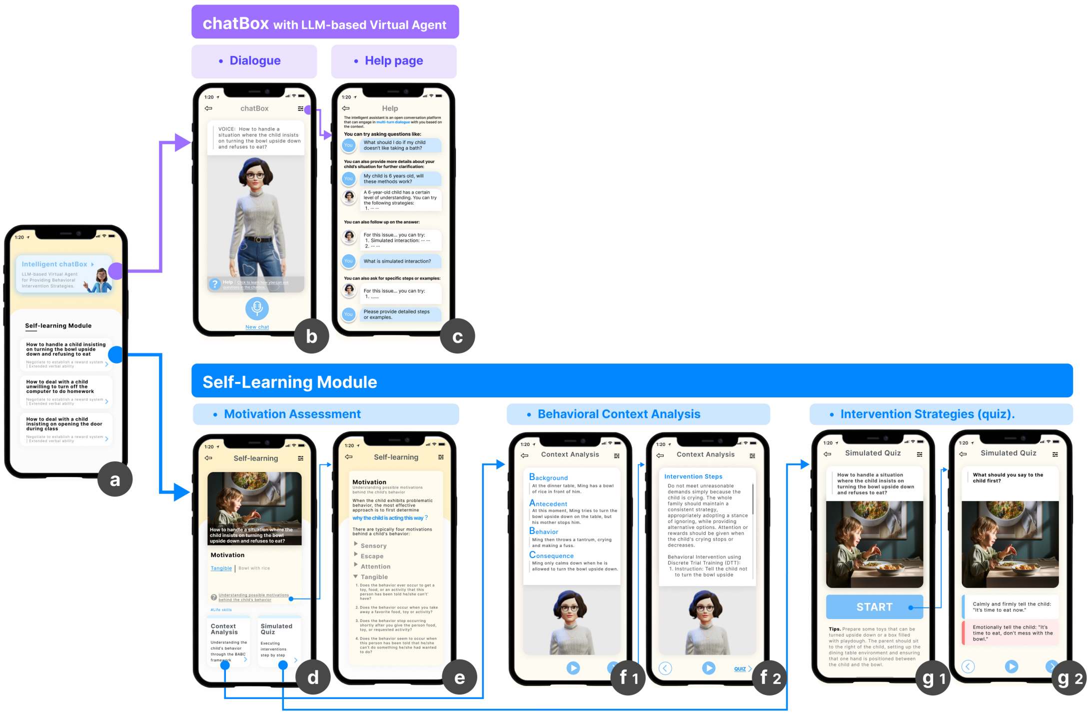

Figure 2. User flow of Glitter.

demonstrated a willingness to learn about intervention knowledge from the SDL module, where they could *"obtain comprehensive knowledge through self-paced or repeated learning and practical case analyses. (P9)"* Participants also perceived the virtual agent as more professional compared to social media or search engines for delivering autism-related content. Participants found the virtual agent's appearance to be friendly and emotionally soothing. Furthermore, participants provided constructive feedback for improving the prototype, suggesting: (1) personalizing the agent's responses, (2) refining and increasing intervention topic and case selection in the SDL module, (3) providing instruction outlines of the BABC analysis steps and intervention strategies, and (4) improving the virtual avatar's facial expressions and body gestures for better affective responses.

3.2. Design Features

Findings from the prior three pilot studies informed the design and iteration of Glitter's key design features. Here, we further report the design and implementation of the finalized application, which was tested in a practical field study with practitioners (Section 4). The Glitter application comprises two modules: the Chatbot, featuring an animated LLM-based virtual agent, and the Self-Directed Learning module.

3.2.1. The Self-directed Learning Module

The Self-directed Learning (SDL) module equips users with a structured roadmap for learning the BABC analysis theory and comprehending the intervention process. The SDL module features 12 cases across 6 topics, i.e., peer interaction, familial dynam-

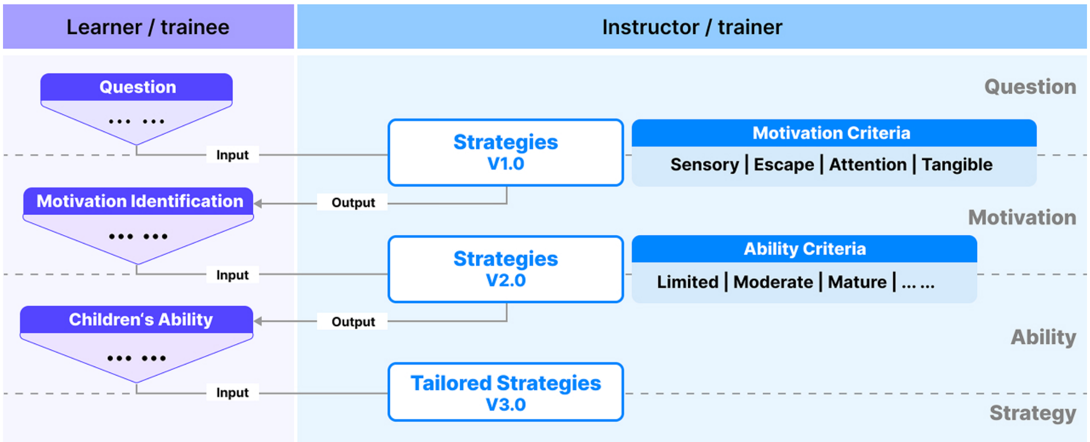

Figure 3. Dialogue framework between learner (or trainee) and instructor (or trainer).

ics, emotion management, life skills, communication, and learning skills. Each case is presented as an interactive unit (Figure 2.a). Each case containes a more detailed analysis (Figure 2.d), with an image and the case summary. Next, users can view an analysis of the child's behavioral motivations and explore the detailed assessment under the Before, Antecedent, Behavior, Consequence (BABC) framework (Figure 2.e). Meanwhile, users can enter the "Context Analysis" unit, view behavior analysis under the BABC framework (Figure 2. f1), and read intervention steps (Figure 2.f2). Users can also access the "Simulated Scenarios Quiz" (Figure 2.g1 and .g2) to initiate a step-by-step test and consolidate their intervention knowledge.

3.2.2. LLM Virtual Agent

Users can also directly ask questions of an LLM-based virtual agent in the ChatBox component using voice inputs (Figure 2.b). The virtual agent replies with both voice and text outputs. Glitter also has a Help page (Figure 2.c) that offers tips on how to ask questions and interact with/prompt the virtual agent.

The retrieval dataset of behavioral issues, practitioners' solutions, and ABA strategies. We constructed a structured dataset to support both retrieval and prompt grounding in Glitter. The dataset comprises 20 common behavioral issues, 51 ABA-based intervention strategies categorized under four behavioral functions (sensory, tangible, escape, and attention), and 334 complete question–and–answer examples. All items were derived from our three pilot studies with practitioners (see Figure 1), which provided real consultation scenarios, intervention discussions, and frequently used strategies. To construct the dataset, two researchers synthesized the pilot study observations into structured entries by: (1) coding behaviors into four functional motivations following FBA guidelines; (2) grouping intervention practices into strategy types; and (3) standardizing practitioner–child scenarios into anonymized Q&A formats. The researchers cross-checked their coding and resolved discrepancies through discussion, resulting in a unified dataset comprising 17 strategy types. To validate the dataset, three licensed practitioners (one BCBA and two senior special-education teachers) independently reviewed the dataset—including behavior categorizations, strategy–function mappings, and Q&A examples. They identified inconsistencies, suggested revisions, and evaluated alignment with established ABA practices. All disagreements were resolved through a consensus review, resulting in the final validated dataset.

The Dialogue Framework. Drawing from established behavioral consultation models and findings from pilot Studies 2 and 3, we extracted a consultation workflow tailored for special education teachers and behavioral therapists. This framework reflects how practitioners typically structure their reflective case analysis and decision-making process based on functional assessment principles.

The consultation workflow includes four key stages (see Figure 3):

- **Question:** Practitioners elicit or articulates the child's presenting problem or challenging behavior.
- **Motivation:** Practitioners identify the underlying motivation behind the behavior.
- **Ability:** Practitioners evaluate the child's individual capabilities relevant to the behavior, such as emotional regulation, communication competence, or developmental profile.
- **Strategy:** Practitioners retrieval, formulate or evaluate a tailored intervention plan with clear objectives and steps.

In practice, these steps may repeat as the conversation evolves. The learner (or trainee) often provides additional information or asks for more detailed guidance, while the instructor (or trainer) uses examples to clarify. We optimized Glitter GPT's responses by setting prompts to limit response length, avoid overly formal language, use examples, break down complex information, and encourage follow-up questions. Detailed prompts are in Appendix E.3.

3.3. Technical Architecture

The Glitter system's technological architecture includes the front-end, the back-end, and the server (Figure 4).

3.3.1. Front-end

The front-end captures audio and video inputs and transmits the data to the server. This part also controls animation clips of the agent's facial expressions and body gestures.

3.3.2. Server

Data Process. The server processes data from the front-end to generate text, audio, and user emotion values, and then sends all generated data to the Unity [6] back-end for further processing. The server initiates the STT process using iFlytek Voice Dictation WebAPI[7]. Then, it runs text-based emotion recognition with a dictionary-based model that maps specific keywords to predefined emotion labels. This model categorizes text into three types: negative, neutral, or positive. For speech emotion analysis, a model based on ResNet34[8] is used. For facial analysis, the server uses the HSEmotion model[9] to classify facial expressions into six categories.

[6]Unity: `https://unity.com/cn`
[7]iFlytek WeBAPI: `https://www.xfyun.cn/`
[8]ResNet34 model for emotion classification: `https://github.com/LaughingPenguin/emotion-classification-resnet34`
[9]HSEmotion model: `https://github.com/HSE-asavchenko/face-emotion-recognition`

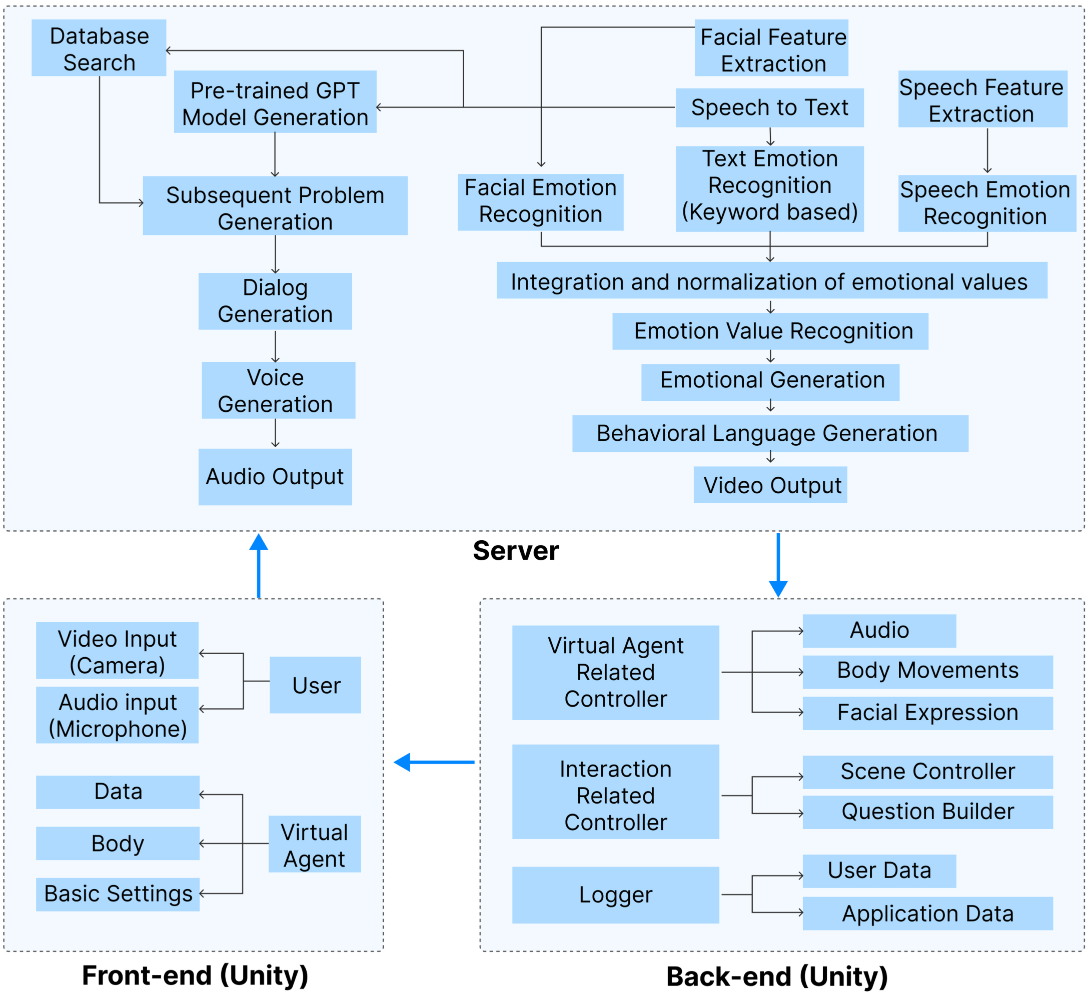

Figure 4. Architecture of Glitter.

To ensure consistency of emotion values, the server normalizes and integrates emotion data. Here, we apply a linear weighting formula. Positive emotions are calculated as $1.2 \cdot \text{happiness} + 1 \cdot \text{surprise} + 0.4 \cdot \text{anger}$, while negative emotions are calculated as $0.8 \cdot \text{fear} + 0.6 \cdot \text{anger} + 1.2 \cdot \text{sadness}$.

These specific coefficients were empirically determined through a calibration pilot test to align the recognition model's output probabilities with the visual intensity requirements of the 3D avatar's blend shapes. For instance, "happiness" was weighted higher (1.2) to ensure positive reinforcement was visually distinct on the avatar's face, whereas "anger" was down-weighted (0.4 in positive contexts) to prevent the agent from appearing inadvertently hostile. Regarding the multimodal fusion, we employ a symmetric 0.5/0.5 linear fusion of text and facial-based emotion scores. Given the heterogeneity of data from different modalities (Pan & Meng, 2024; Wang et al., 2023), linear fusion provides a robust, low-latency baseline that avoids the overfitting risks of complex non-linear fusion models in real-time interaction scenarios.

In addition, the server is also responsible for dialogue generation and voice generation. After generating the response from ChatGPT-4o, the server converts the text into speech using the iFlytek Speech Synthesis API (model v2) with a Mandarin voice ("xiaoyan") and default prosody settings[10] to produce an audio clip for the virtual agent to "speak" the reply. A complete list of API configurations is provided in Appendix F.

The GPT Agents and Prompts for the Conversations. The system utilizes two GPTs: the Dialogue Management GPT and the Response GPT. The first GPT handles multi-round conversations to ensure continuity, while the second GPT generates responses. The Response GPT is supported by a pre-prompted database (described later) of FAQs and strategies related to autism interventions. The prompts are demonstrated in Appendix E (originally in Chinese). The pre-prompted database includes strategies for managing common behaviors, such as tantrums, communication difficulties, and sensory sensitivities. When a user's query is in the database, the GPT retrieves and outputs the predefined response. If a user's query does not match any entry in the pre-prompted database, the Response GPT automatically switches to a generation mode. In this case, it produces an answer by combining (1) the structured information collected by the Dialogue Management GPT during the multi-round conversation, and (2) its own language-model reasoning. The two-GPT architecture ensures that the generated response still aligns with the practitioner-centered reasoning framework described earlier. In other words, the predefined database enables high-quality retrieval for common scenarios, while the generative mode provides flexible support for novel or less frequent practitioner questions.

3.3.3. Back-end

The back-end integrates and processes data from various input sources and controls the entire UI and virtual agent. After extensive exploration, a final weighting method of $final\_emotion\_value = 0.5 \times facial\_value + 0.5 \times text\_value$ was adopted to calculate the final emotion value. The relative reliability of facial and text cues is highly contingent upon different environments. Recent research indicates that effectively integrating multimodal information remains a major challenge due to the heterogeneity of data across modalities (Pan & Meng, 2024; Wang et al., 2023). In this regard, we employ a symmetric 0.5/0.5 linear fusion of text and facial based emotion scores as a conservative yet robust default approach. This approach refrains from presuming a

[10] `https://global.xfyun.cn/doc/tts/online_tts/API.html#description-of-the-interface`

globally dominant modality and remains straightforward to interpret. Moreover, although voice emotion has been collected, it has been found to have low accuracy and is omitted in the final emotion value calculation. Afterward, the back-end determines the appropriate animations of the virtual agent to achieve affective responses. The Virtual Agent Related Controller selects animations based on Table G1 in Appendix G.

4. Methodology

We conducted a one-week, in-the-wild field study with practitioners to examine how Glitter fits into their everyday behavioral consultation workflows and how they perceive its usefulness, real-world adoption, and the support provided by the self-learning module and LLM agent.

4.1. Participants

We recruited practitioners via purposive sampling through professional special education networks and WeChat groups. We established the following inclusion criteria: (1) currently employed as a special education teacher, behavioral analyst, or therapist; (2) possess at least one year of experience working with autistic children with a focus on behavior management; (3) fluent in Mandarin Chinese; and (4) own a mobile device capable of installing the prototype.

We initially recruited 18 practitioners. To ensure data validity, we applied exclusion criteria based on engagement logs. Participants were excluded if they used the app for fewer than five days or accumulated less than ten minutes of total active usage. This threshold was set to filter out "novelty bias" or superficial interactions unsuitable for evaluation (Barnett & Lewis, 1994; Florencio & Herley, 2007; Lazar, Feng, & Hochheiser, 2017; O'Brien & Toms, 2008). The gender composition (91% female) is consistent with the female-dominated workforce in China's special education and rehabilitation sector (World Health Organization, 2018).

The final sample consisted of 11 participants (10 female, 1 male; aged 20–31). Detailed demographics are shown in Table 2. As a result, five participants were excluded based on the engagement criteria, one withdrew because the app did not recognize her dialect, and another opted out because the app did not provide relevant content for her speech therapy practice. At the end of the study, 11 participants (10 female, aged 20-31) from 8 different cities completed the study. Five are special education teachers in the primary and preschool stages, who worked in special education schools or institutions. Four are rehabilitation therapists for autistic children, and the remaining two are education technology teachers who work in special education schools, as well as shadow teachers (who serve as private tutors for one autistic child in primary school). This research study was approved by the *Duke Kunshan University* IRB (Approval No. FWA00021580). All participants received monetary compensation for their time.

4.2. Procedures

After enrollment, we informed them of the objectives, procedures, and all relevant details, and then obtained their consent to participate. Participants were encouraged to use the Glitter app as often as possible—particularly when they needed support or wanted to reinforce their knowledge about the BABC approach (and related information) in the self-learning module. We did not require participants to use Glitter at specific times or in certain situations, as our goal was to collect data about actual use case scenarios and insights from their daily work and life. We also informed the participants that they could withdraw from the study at any time. The entire study was conducted online via video conferencing software.

Pre-Study Tutorial and Setup. In the pre-study session, we first sent the installation link via TestFlight[11] (iOS) or APK package (Android) to the participants and asked them to install the software on their own mobile devices. We collected their demographic information, work-related experiences, and general use of AI applications in their daily work with autistic children. We walked the participants through the app using screen-sharing to ensure they understood how to use it. We next sent an instruction manual and a demo video so that participants could review how to use it if they encountered any problems or had questions during the study. We invited participants to use Glitter as often as needed in their daily lives and work to receive practical feedback and insights from the field (i.e., compared to a lab study). We also asked the participants to take notes or share messages containing their feedback, reflections and thoughts every time they used it.

One-Week Field Study. In the one-week study, we tracked participants' in-app usage and did not reach out to them unless they had any questions or shared notes with us. We provided participants with a few conversation examples to illustrate how users could interact with the agent, whose content was derived from our prior pilot studies with caregivers and practitioners, as described in Section 3.1. See the conversation examples in Appendix A.

[11] `https://developer.apple.com/testflight/`

Table 2. Demographics of participants in the field test.

ID	Age/Gen	Education/Major	Work Experience w/Autistic Children	City
P1	25/F	Bachelor/ Rehabilitation Therapy	Training Institution (2 years): One-on-one Training, Speech Therapy	Suzhou, Jiangsu
P2	23/F	Bachelor/ Special Education	Special Education School (1 year): Life Skills, Crafts, and Drawing	Yunfu, Guangdong
P3	31/F	Bachelor/ Special Education, Preschool Education	Training Institution (1 year): One-on-one and Group Training	Beijing
P4	24/F	Associate/ Rehabilitation Therapy Technology	Training Institution (2 years): Speech Therapy	Suzhou, Jiangsu
P5	27/F	Master/ Educational Technology	Special Education School (2 years): Information Technology, Life Skills, Crafts and Drawing	Shenzhen, Guangdong
P6	20/F	Bachelor/Special Education	Research Institute (1.5 years): VB-MAPP Application	Zhanjiang, Guangdong
P7	25/F	Bachelor/ Rehabilitation Therapy Technology	Training Institution (1 year): One-on-one Training/Integration	Suzhou, Jiangsu
P8	23/F	Bachelor/ Rehabilitation Therapy	Training Institution (1 year): One-on-one Training, Speech Therapy	Suzhou, Jiangsu
P9	29/M	Bachelor/ Mechanical Design, Manufacturing and Automation	Training Institution, Mainstream School (5 years): Social Games, Sensory Integration, Shadow Teacher	Dongguan, Guangdong
P10	20/F	Bachelor/ Special Education	Private Tutor (1 year): Life Literacy, Life Math, Life Skills, and Rhythm	Guangzhou, Guangdong
P11	25/F	Bachelor/ Special Education	Training Institution, Special Education School (5 years): One-on-one Training, Life Math	Shantou, Guangdong

Post-Study Evaluation. At the conclusion of the study, we invited participants to a post-study online debrief session and asked them to reflect on their experiences in a semi-structured interview. Participants also completed a questionnaire with a self-reported usability scale and other user experience ratings.

4.3. Measurements

We collected participants' perceptions, experiences, and feedback using Glitter through a usability evaluation questionnaire, users' application logs, and semi-structured interviews.

4.3.1. Self-reported Usability and User Experience Questionnaires

We collected participants' self-reported feedback using a five-point Likert scale with questions adapted from the Questionnaire for User Interface Satisfaction (QUIS) (Chin, Diehl, & Norman, 1988) and the System Usability Scale (SUS) (Brooke, 1996). Users were asked to rate the Glitter system on its general usability (Q1-3), overall user experience (Q4-7), emotional support and empathy (Q8-9), personalization (Q10-11), security and privacy (Q12-13), and overall satisfaction (Q14-15). Meanwhile, participants also filled out the SUS scale (Q16-25). See the full question list in Table 4.

4.3.2. App Logs

We collected users' app usage data and stored it on our private data server, including users' login and logout timestamps, conversation content with the LLM agent, and the time users spent in Glitter.

4.3.3. Expert-based Evaluation of LLM Responses

To obtain an objective assessment of the LLM agent's generated responses, we conducted an expert-based content evaluation. Two senior special education practitioners (each with over ten years of professional experience) independently reviewed 154 practitioner–agent Q&A instances collected during the field study. Each response was rated along three dimensions: overall response quality, domain knowledge accuracy, and practical actionability, using 10-point Likert-type scales. Experts could also provide open-ended comments to note major issues or concerns.

4.3.4. Semi-structured Interviews

In the semi-structured interviews, we collected participants' feedback on the SDL module's content and the integration of the BABC framework, as well as their evaluation of the learning effectiveness. We also asked participants to evaluate the LLM agent's answers, use scenarios, and usefulness, as well as provide feedback on the agent's multimodal attributes. See the interview guide in Appendix D.

4.4. Data Analysis

4.4.1. Quantitative Data Analysis

For the self-reported usability and experiences questions, we conducted descriptive analyses and reported the mean and SD values. For the time spent in the app, we calculated the total and average time spent on each feature by users.

4.4.2. Qualitative Data Analysis

We analyzed two qualitative data sources: conversational logs and semi-structured interviews. For the conversational logs, two researchers systematically analyzed and categorized interaction patterns (e.g., question types, query processes). For the interview transcripts, we employed a hybrid thematic analysis approach (Fereday & Muir-Cochrane, 2006), which combined both deductive (top-down) and inductive (bottom-up) coding, following the principles of Braun and Clarke (2006) (Braun & Clarke, 2006).

To ensure analytic transparency and rigor, two researchers first established a set of deductive codes based on our interview guide. They then independently coded all transcripts, applying the initial codes while also inductively generating new codes for emergent data. To ensure reliability, the researchers held iterative meetings to compare codes, discuss discrepancies, and resolve them through negotiated consensus (McDonald, Schoenebeck, & Forte, 2019). This process resulted in a unified codebook that was then organized into the final themes (see Section 5.3). To assess inter-rater reliability, two researchers independently coded a randomly selected subset of the interview transcripts (3, 27.27%) using the final codebook, and Cohen's kappa was calculated to quantify coding agreement. Illustrative excerpts from the interviews are provided in the findings to substantiate the generated themes.

4.4.3. Expert-based Evaluation Analysis

Expert ratings were analyzed descriptively by calculating the mean and standard deviation for each evaluation dimension across all reviewed responses. Given the exploratory nature of this study, these ratings were used to characterize the overall quality and practicality of the generated content rather than to support inferential hypothesis testing. Inter-rater reliability (IRR) was assessed using a two-way random-effects intraclass correlation coefficient (ICC).

4.5. Ethical Considerations

Three senior coauthors have 3-5 years experience in working with autistic children and relevant stakeholders, who hold an inclusive perspective that understands ASD as a form of neurodiversity rather than a disorder. Two senior authors carefully guided the entire project to make sure we used respectful language, protected participants' data, and provided trustworthy materials throughout the entire research process, including data collection, design feature extraction and prototype design, user studies, and findings analysis. In Glitter, we did not use any specific autism-related words (including but not limited to "autism, disorder, disease, disabled," and so on) in describing the content to avoid triggering the potential social stigma of caregivers or practitioners. All study data were stored on secure, university-managed servers with access restricted to authorized research team members. Data transmission was encrypted. In accordance with institutional policy, raw research data will be retained for five years and used only for research purposes related to this study.

5. Results

5.1. Quantitative Findings of System Usability and Users' Engagement

5.1.1. Self-reported Usability Scales

In the post-study self-reported questionnaire, a descriptive analysis suggested that practitioners had medium-to-high ratings, with most item means falling between 3 and 5, as illustrated in Figure 5. Findings from the SUS scale (Mean = 83.64, SD = 9.71) also showed that practitioners were generally satisfied with the Glitter app. Such findings were in accordance with the interview results, as indicated in later subsections of qualitative findings. See Table 3 for a full summary of usability and engagement metrics.

5.1.2. Usage Time Across Modules

Practitioners used the SDL module for a total of 490.25 minutes, with an average of 6.37 ($SD = 4.46$) minutes and 15.97 ($SD = 9.26$) interactions per day. In comparison, they spent 717.50 minutes using the LLM agent, averaging 9.32 ($SD = 3.78$) minutes and 4.77 ($SD = 1.18$) interactions per day. To examine whether these differences reflected distinct engagement patterns, we conducted paired t-tests on daily usage. The difference in daily time spent across modules did not reach statistical significance ($t(10) = 1.94$, $p = .081$), whereas the SDL module was accessed significantly more frequently ($t(10) = -4.27$, $p = .002$). These results suggest that practitioners engaged with the LLM agent in longer sessions, while the SDL module was used in shorter but more frequent segments.

5.1.3. Expert-based Assessment of Response Quality

The expert-based evaluation indicated that the LLM-generated responses were generally perceived as useful and applicable. Across all evaluated responses, ratings were moderate to high on overall quality ($Mean = 7.26, SD = 2.72$), knowledge accuracy ($Mean = 7.10, SD = 2.70$), and actionability ($Mean = 6.95, SD = 2.74$). Inter-rater agreement was fair across the three dimensions (ICC range = 0.25–0.29, $p < .05$). Descriptive patterns suggested systematic differences in rating tendencies between the two experts, with one applying a stricter scoring standard and the other showing more conservative variance. This variability reflects differences in professional judgment criteria rather than random disagreement. Overall, the expert assessment provides converging, objective support for practitioners' qualitative perceptions, while also highlighting the inherent subjectivity involved in evaluating intervention content.

5.2. Conversation Logs with the LLM Agent

We analyzed practitioners' conversation process and content with the LLM agent and their conversation prompting strategies in obtaining satisfied and ideal answers.

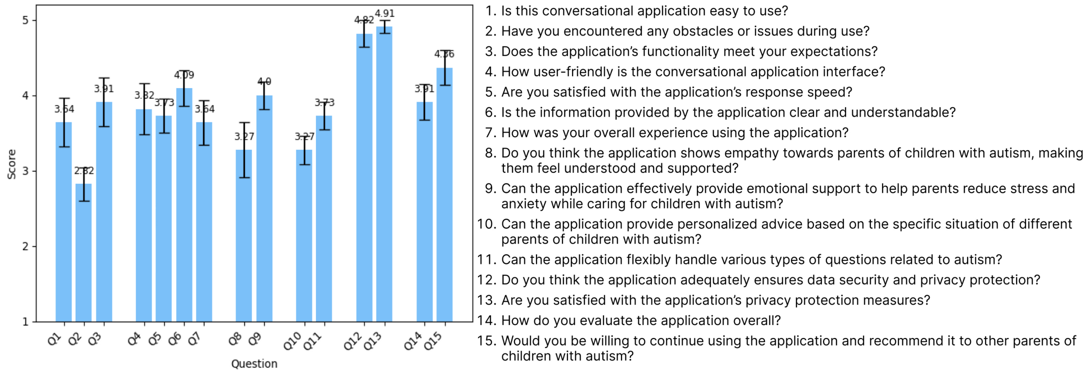

Figure 5. Mean values and standard errors for all questionnaire items.

Table 3.: Summary of Key Usage Metrics and System Engagement (7-Day Study)

Metric Category	Measure	Value (M, SD)
Overall System Usability	SUS Score [0-100]	83.64 (SD = 9.71)
Module Engagement (Daily)	Time spent (LLM)	9.32 (SD = 3.78) min/day
	Time spent (SDL)	6.37 (SD = 4.46) min/day
	t(10)=1.94, p=.081	
	Frequency (LLM)	4.77 (SD = 1.18) times/day
	Frequency (SDL)	15.97 (SD = 9.26) times/day
	t(10)=-4.27, p=.002	
Module Engagement (7 days)	Total time (LLM)	717.50 min
	Total time (SDL)	490.25 min
Overall Interaction Logs	Total App usage time	1622.68 min
	Avg daily App usage	231.81 min
	Multi-turn dialogues	68 sessions
	Single-turn queries	314 queries
	Unique inquiry topics	65 topics
LLM Top Inquiry Categories	Learning Skills	22 (33.8%)
	Behavioral Challenges	14 (21.5%)
	Communication Proficiency	11 (16.9%)

Table 4. Inquiry question types and examples.

Topic	Freq.	Example
Attention	5	*What should I do if the child has poor attention and keeps looking around during class? (P8)*
Behavioral Challenges	14	*What should I do if the child likes to run around? (P2)*
Communication Proficiency	11	*How can I help a child with multiple answers to a question express themselves proactively? (P7)*
Diagnostic Queries	1	*What is autism? (P9)*
Emotion Management	2	*How can I intervene for a 6-year-old autistic girl with self-stimulatory behaviors and emotional distress? (P3)*
Learning Skills	22	*How can I help a child transition from matching physical objects to matching cards? (P1)*
Life Skills	5	*The child is already 6 years old but still refuses to sleep separately from adults. (P9)*
Peer Interaction	2	*How can I encourage the child to be interested in playing with other children? (P9)*
Social Interaction	3	*The child ignores what others say. (P10)*

5.2.1. Inquiry Question Types

In total, practitioners initiated 68 multi-round conversations with the LLM agent, covering 314 single-round Q&A data items. Some conversations covered multiple topics, as users did not always start a new session when switching subjects. We categorized the query questions and identified 65 topics, falling under nine intervention/skills types, such as Attention, Behavioral Challenges, Communication Proficiency, and others. The most frequently queried topics were Learning Skills (N = 22), Behavioral Challenges (N = 14), and Communication Proficiency (N = 11). See Table 4 for details.

5.2.2. Coverage of the Retrieval Dataset in Field Queries

To better understand how the current retrieval dataset aligns with real practitioner use, we examined whether field-study queries fell within the scope of the 20 predefined behavioral-issue topics included in the dataset. Each query group was coded as either (1) directly related to these behavioral-issue topics, (2) adjacent to the framework but not explicitly represented among the 20 issues (e.g., attention training, social participation, generalization), or (3) outside the current dataset scope (e.g., language training, academic instruction, concept learning). Across 154 query groups, 54 (35.06%) aligned directly with the dataset's behavioral-issue topics, 31 (20.13%) fell into adjacent ABA-related areas, and 69 (44.81%) concerned domains not currently covered by the dataset (e.g., language training). These results contextualize the system's current coverage boundaries and motivate future work on expanding the dataset and conducting a more fine-grained coverage analysis.

5.2.3. Conversation Process and Prompting Strategies

We analyzed the conversation logs and observed that practitioners adopted three main prompting strategies.

Single Inquiry for Immediate Answer Practitioners frequently ask a question to receive a direct and effective response in a one-round conversation with the agent. This was commonly seen for clear-cut topics, e.g., *"What is floor time?"* (P4), *"What is autism?"* (P9), and *"How to perform discrete thinking and teaching?"* (P6).

Rephrasing Query Language for Better Answers When practitioners could not receive satisfactory answers for more complex issues, they reworded their questions by adding more details or rephrasing their sentence structures. For example, P5 elaborated on the same question for four attempts to get the best answer: (1) *"What to do if a child hits others?"*, (2) *"How to deal with a 6-year-old autistic boy who attacks others?"*, (3) *"How to handle a 6-year-old autistic boy pushing, hitting, and cutting in line?"*, (4) *"What to do if a 5-year-old autistic boy pushes, cuts in line, and grabs toys during play?"*. Similarly, P7 also rephrased the problem: (1) *"The child only talks to familiar people and is silent with strangers."*, (2) *"The child is happy in familiar environments but goes silent in new ones. What can be done?"* Rephrasing helped participants get more specific answers.

Progressive Queries for Iterative Solutions In *"How to"* type questions, practitioners often refine their queries to gain more detailed or personalized intervention strategies. For example,

(1) P1, *"What to do if a child doesn't follow instructions?"* The agent: *"... Use reinforcements like toys... "*.
(2) P1, *"Even with rewards, the child remains unmotivated and disobedient."* The agent: *"... Provide more detail on the child's behavior..."*.
(3) P1, *"The child gets distracted during the process."* The agent: *"... Try using preferred visual cues..."*
(4) P1, *"Even with prompts, the child shows little interest..."* The agent, *"... Simplify steps and offer immediate rewards..."*.

In this case (and other similar conversations), P1's initial question was too broad, so she received a general strategy. However, as the agent prompted P1 for more details through multiple rounds of conversation, the agent suggested a more tailored intervention, specifically, adjusting the timing of rewards.

5.3. Qualitative Findings from Semi-structured Interviews

This section contains thematic analysis results of practitioners' use case scenarios and perceived usefulness of Glitter in the field study (RQ1), the potential influence of the LLM agent's multi-modality attributes to practitioners' information disclosure and acquisition (RQ2), thoughts on opportunities and challenges in adopting Glitter in practical scenarios (RQ3), and feedback on how other relevant stakeholders might benefit from the LLM agent (RQ3). The key themes emerging from our analysis of these questions are illustrated by the thematic structure presented in Table 5.The inter-rater reliability for the qualitative coding was substantial (Cohen's $\kappa = 0.76$), indicating satisfactory agreement between coders.

5.3.1. Practitioners' Perceptions of the Self-directed Learning Module

Most practitioners (8, 72.7%) commented that the self-directed learning (SDL) module provided a clear, easy-to-understand, and practical learning experience, which helped them become more familiar with or strengthen their knowledge of the BABC analysis process under the ABA framework. For instance, P2 evaluated the SDL module as "systematic and structural",

> *"The internal learning logic [i.e., the integration of the BABC principles within intervention strategies] of the self-directed learning module is very clear, which starts with explaining behavior issues, providing overviews of intervention strategies, and then practical steps."(P2)*

Table 5.: The themes derived from the thematic analysis.

Main Theme & Sub-theme
T1: Perceived Value as "Just-in-Time" & "Just-in-Case" Tool
- On-demand Intervention Support and Inspiration
- Bridging Gaps in Professional Knowledge (Especially for Specific Groups)
- Low-Cost Emotional and Informational Support for Caregivers
T2: Content Quality: Professionalism vs. Lack of Diversity
- Strengths in Core ABA vs. Gaps in Specific Domains
- Balancing Professional Jargon with User Accessibility
- Lack of Diversity in Case Examples
T3: Dual-Feature Design: Complementary and Disconnected
- Complementary Use Cases (Immediate Answers vs. Systematic Browsing)
- Strong Need for Feature Integration
T4: Usability and Engineering Challenges
- Missing Core Functionalities (Text Input and History)
- Interaction Friction and Bugs
T5: Agent's Role: Professional Symbol vs. Emotive Reader
- Appearance: Professionalism Builds Trust, but Style is Debated
- Voice as the Key Perceptual Channel (but Lacking Emotion)
- Function over Form (Content Prioritized over VA)
T6: Risks, Challenges, and Ethics
- Risk of Over-reliance and Misinformation
- Privacy Perceptions (Anonymity as a Benefit, Voice as a Risk)
- Limitations of LLM Responses

Around half of the practitioners (5, 45.5%) thought that the quiz questions supported them in better comprehending and memorizing the practical implementation of the BABC framework in different use cases. P11 reflected that even though she was experienced in the ABA intervention framework, she sometimes still did not select the correct options in the quiz testing stage. Two practitioners spoke highly about the practicality of "group skills" since the SDL module offers strategies for addressing behavior issues, which P4 thought aligns with the goal of integrating children and meeting key home intervention goals. To conclude, most practitioners viewed the SDL module as an effective learning platform and adopted it for recapturing and refreshing the intervention framework.

In addition, practitioners (6, 54.5%) also provided improvement suggestions for the SDL module. P3 mentioned that the SDL module could include more behavioral analysis and intervention methods besides the BABC process and the DTT solutions, which could allow users to try multiple approaches until they find the most effective interventions for children's behavioral issues. Although the SDL module addressed understanding children's personal characteristics (e.g., their language abilities, motivation types, and social skills), P1 mentioned that its materials should also consider the diversity of children's behavioral motivations, which may be joint and complex.

5.3.2. Practitioners' Use Case Scenarios of Adopting an LLM Agent for Behavioral Interventions

We discovered that practitioners adopted the LLM agent under varying practical situations, depending on their teaching/intervention experiences and purposes. The most frequently reported use cases fell into the following categories, in which practitioners found the conversations useful and described them as sources of guidance or inspiration:

- **Inquire & Resolve.** Inquiring for detailed intervention content and overall conceptual framework or progressive steps (i.e., high-level strategies) that practitioners did not have an answer for or know what to do (7, 63.6%);
- **Compare & Improve.** Comparing LLM's intervention suggestions with their current solutions to select the best one, or merge LLM interventions to formulate a new one (7, 63.6%);
- **Prepare & Conclude.** Preparing teaching materials before classes or summarizing intervention strategies after classes, which practitioners use as a source of information and supplementary materials (6, 54.5%)

Four experienced practitioners stated that they received useful inspiration from the LLM agent when they encountered unsolvable issues with children's behavioral problems and other non-behavioral intervention challenges. Practitioners believed the agent served as a reference for solutions and facilitated the integration of suggested interventions into their existing approaches. P9 gave an example of the **Inquire & Resolve** scenario from her one-on-one class,

> *"My student couldn't understand 'countdown,' so I consulted the agent and applied its suggestion of using tokens and removing items one by one. To my surprise, this method worked well. Although I had used physical objects before, they were not used in this specific counting process. I plan to try and expand this physical teaching method in other contexts." (P9)*

Four practitioners (36.4%) also mentioned they used the app to inquire about the progressive steps or overall conceptual framework, where they used it as a practical guidebook. The agent provided guidance for them to assemble their personalized and comprehensive intervention process without "having to read a domain-specific textbook" (P11). P2 shared one of her use case scenarios, which also fell under the **Inquire & Resolve** scenario.

> *"I mainly used the App to generate intervention steps according to the ABA framework... Then, I adopt these steps to formulate my own practical steps. As for the specific intervention steps, I prefer to rely on my own knowledge and past experiences···" (P2)*

For practitioners' **Compare & Improve** type of use case scenarios and motivations, they usually already had their own intervention strategies and content, but they still chatted with the LLM agent. They compared the agent's suggestions with their own approaches to identify a better one or improve their own interventions. Take P8's experience,

> *"I was working one-on-one with a child with low comprehension, making communication tough. The app offered me practical steps based on DTT. I also adopted DTT, but the agent's guidance made me realize I hadn't been giving enough reinforcement or enough time for the child to respond. After adjustment, the child's performance improved." (P8)*

In another **Compare & Improve** scenario, some practitioners also said that when they received parents' inquiries about intervention strategies for children's behavior issues, they would first try to talk with the LLM agent for intervention suggestions and then compare with his/her existing approach. The practitioners then compared the agent-suggested approaches with the strategies they had in mind, and identified the better one or improved the approach in their mind through careful testing with children. P7 said she would test the new approach in her classroom with the specific child and observe the practical outcomes the next day; she only shared the successful ones with the parents who asked those questions.

The third most frequently mentioned use case scenario was **Prepare & Conclude**, referring to preparing upcoming lessons and concluding lesson plans and teaching materials for behavioral intervention classes. Some practitioners mentioned that they would chat with the agent to gather information about their class's general topics during the lesson preparation stage, treating the LLM agent as a reliable source of information, alongside materials from books or their training notes. Several practitioners also mentioned that they had chatted with the agent about some interventions or behavioral issues that had already occurred in prior classes, as they wanted to quickly confirm their intervention strategies or review specific and easy-to-forget knowledge and intervention techniques.

5.3.3. The Multi-modality Attributes of the LLM Agent: Visual Appearance, Speech, and Affective Responses

Besides the LLM-generated conversations, practitioners also commented on how the agent's appearance, speeche, and affective responses (i.e., facial expressions and body language) affected their overall user experience, information disclosure, expression, and acquisition.

The LLM agent's speech provided memorable and impressive experiences. Practitioners (7, 63.6%) appreciated the voice feature of the agent, and they thought that the agent's speech highly increased their sense of talking to "someone", like chatting with a person in real life. For instance, P7 responded that voice was one of the most important embodied features when talking to people, so even if she knew it was a cartoony avatar, she felt a sense of social presence of a human being. P8 and P9 reported that their most memorable features of the agent were her speech, then her visual style and hand gestures; they paid the least attention to the agent's facial expressions. Sometimes, the agent responded with a large volume of text that required time for reading and comprehension. Therefore, P6 suggested generating prosody information ("like emphasis and pauses") in the agent's voice to attract users' attention and improve their efficiency in comprehending the conversations.

The LLM agent's visual appearances were embodied, professional, and friendly, but did not affect practitioners' overall information expression and acquisition. Most practitioners responded that they had not paid much attention to the visual appearances of the agent and over half of the practitioners (7, 63.6%) reported that the agent's visual appearance did not affect their willingness to disclose questions, private information, or desires of expression, partially because that they put the urgency of acquiring information and knowledge for behavioral interventions at top priority (P8) and they knew that the agent was not "real" (P11). For instance, P7 said she received most of the information from the text/speech and had to comprehend it, which took up her full cognitive capacity.

However, practitioners still felt the presence of a 3D virtual avatar was needed, offering a sense of physical existence and embodiment of an "imaginative but real" character/figure (P8). When asked about their specific feedback for visual attributes, most said that they liked the visual style and perceived the look and feel of this agent to be "professional," "warm," and "friendly" (4, 36.4%). P11 also said that she felt the virtual avatar's professional dress and overall visual presence were critical reasons that made her trust in the conversation content with the agent.

The LLM agent's affective responses were not well-perceived and not critical in comprehending information. Practitioners noticed that the agent was generally in a good mood and "always smiling", but most did not really notice obvious affective changes occurring in the virtual avatar's facial expressions or body gestures during conversations. As mentioned earlier, more than half of practitioners (7, 63.6%) responded that they paid more attention to the conversation content in speech and texts. One practitioner responded that the subtlety of facial expressions made them difficult to detect, leading participants to perceive the agent as "constantly smiling" (P8). Although the system triggers animations based on user states detected (Section 3.3), we did not independently validate the reliability of this sensing pipeline in real-world use; thus, we refrain from attributing the perceived lack of affective change to specific technical factors. Instead, we interpret this as a perception-level finding: affective cues were often not salient enough to be noticed during task-focused interactions. In contrast, three practitioners stated that they hoped to feel "being cared" for by the agent during their conversations, and mentioned the essential role of facial expressions in supporting them in perceiving the agent's empathetic attitudes. For instance, P1 said, *"...her calm expression gives me a sense of comfort. But if her expression never changes, it makes me doubt whether she has any empathy."*.

5.3.4. Practitioners' Perceptions of the LLM Agent's Potential Benefits in Behavioral Interventions for Other Stakeholders of Autistic Children.

Practitioners also believed that the intervention suggestions provided by the agent could benefit other stakeholders of autistic children, including teachers working with large classes at K-12 general schools or in under-resourced or remote regions, college or university students majoring in special education, and especially parents or other caregivers of autistic children.

Teachers from special education schools and general schools. Three practitioners who worked in special education schools thought the agent helped them refresh the ABA framework and also other behavioral intervention knowledge. Unlike practitioners at special education institutions or centers whose work mainly focuses on interventions and therapies, practitioners at special education schools have more responsibilities in managing a larger group of students. They often face complex teaching tasks, heavier workloads, and diverse cases that require tailored responses, which limits their focus on behavioral intervention content and techniques.

Teachers from low-socioeconomic/remote regions. Though none of our practitioners were from low-socioeconomic or remote areas, some noted that teachers in these regions often have fewer resources and less access to training or peer support. The LLM agent could help them access information and guidance on behavioral interventions for children with special education needs.

College or university students majoring in special education and behavioral therapies for autistic children. Some practitioners, e.g., P6 and P10, reflected on their time as college students majoring in special education. Although they learned strategies and concepts and observed other teachers' classes, the beginner practitioners usually lacked working experience. This makes it hard for them to recognize behavior issues and understand how interventions vary for each autistic child. They felt conversations with the agent provided practical strategies for managing behavior issues, which P10 found especially helpful during her transition from student to teacher.

Caregivers (i.e., parents) of autistic children. All practitioners agreed that Glitter would be a valuable tool for caregivers of autistic children to learn the application and integration of ABA behavioral analysis and intervention strategies. They noted that while they taught for just a few hours a week, caregivers usually spent the most time with the children, often encountering the most challenging behaviors. Furthermore, practitioners noted that caregivers had limited resources and technological support in acquiring knowledge or information for individualized cases. Parents often feel desperate when interventions fail and rely on teachers for guidance, which can be time-consuming, especially for children in general or special education schools. Practitioners believed the LLM agent could offer more accessible and synchronous instructions to improve parents' confidence and reduce stress in family interventions.

5.3.5. Practitioners' Challenges and Concerns in Using the LLM Agent

Besides the general positive feedback and potential for use by other stakeholders, most practitioners (10, 90.9%) also shared their challenges and concerns that occasionally happened during the one-week field study.

Practitioners perceived writing good questions (prompts) to query effective answers from the LLM agent as being challenging. As indicated in Section 5.2.3, practitioners explored multiple queries and prompting strategies to obtain the best answers. Some practitioners were unfamiliar with prompting the LLM multiple times to receive the best answers. If the LLM agent did not offer a satisfactory answer in the first round of chat, participants were unsure how to revise and iterate their prompting questions to progressively receive the expected answers. Therefore, several (4, 36.4%) reflected that they thought the quality of the answers was sometimes unstable, so they had to try different ways of asking questions to receive the most desired replies. Two mentioned that if they did not receive a good-quality answer in the first round, they felt the agent was incapable of answering that specific question, so they gave up on that question and tried to ask other questions instead.

Practitioners had concerns about the LLM agent's language expression: terms, styles, and specific nuances. A majority of participants (72.7%) expressed concerns about the LLM agent's language expression approach. While practitioners had no trouble understanding the content due to their training, they noted that some academic terms might confuse non-professional users, such as parents, early-career teachers, and non-professionals. Although practitioners appreciated the detailed intervention content and steps, they found certain nuanced wording to be somewhat general, which might create confusion in understanding the specific meanings, such as the ways of referencing subjects, naming of objects or events, and placement of characters in a sentence. They emphasized that precise language is crucial for clear communication and effective implementation of interventions.

Practitioners' concerns about parents and other non-professional users' potential limitations in judging the appropriateness of interventions or transferring knowledge to practice. Four practitioners expressed their concerns about several problems parents (and other non-professional stakeholders) could encounter if they were the users of this system. The primary concerns included: (1) the limited diversity of intervention strategies might constrain parents in formulating flexible intervention approaches for their children; (2) parents could not sufficiently transfer intervention strategies learned from the agent and successfully deliver them to their children in various scenarios; and (3) parents might over-trust the LLM agent and be overly dependent on everything it offered, which might create oversights, mistakes or even risky consequences.

Practitioners expressed preferences for exploring other intervention strategies and techniques beyond the ABA framework, BABC, or DTT techniques. Practitioners agreed that ABA is a key guiding principle in their work but expressed interest in exploring other intervention approaches. As P3 noted, *"ABA is common, but not always suitable; some interventions require motor development strategies from doctors."* Furthermore, since the LLM agent's response content and structure were primarily referenced from the dataset we constructed, P3 and P8 noticed that the LLM agent's responses tended to become homogeneous in a few conversations under certain scenarios. When we prompted and tried to confine LLM responses, certain keywords had higher weights or occurrences in the dataset, which were overly emphasized and thus reduced the diversity of answers. The unexpected repetition in the response structure limited practitioners' further explorations and continuing conversations with the LLM agent. For instance, DTT appeared frequently in the "how to" questions, when practitioners seek support in intervention solutions.

Practitioners were generally satisfied with Glitter's privacy and security settings. Interestingly, most practitioners (10, 90.9%) felt comfortable sharing their personal cases in the classroom, including the characteristics and situational descriptions of their children, and their teaching materials and lesson plans. Meanwhile, most practitioners (10, 90.9%) mentioned that they did not have any privacy or security concerns, mainly because *"the app does not require users to provide much personal information, such as the child's name, age, diagnosis, or the parents' identity"* (P1). She responded that anonymous inquiries reduced her concerns about revealing the child's condition during a conversation with the agent, providing a sense of security. Another participant (P2) mentioned that *"there are no labels like 'autism' or special markers in the interface. It feels like a regular app, which is great."*

6. Discussion

Drawing on findings from our pilot studies and one-week field deployment, we reflect on our research questions and discuss implications for designing practitioner-facing LLM agents in behavioral intervention contexts.

6.1. Facilitating Dynamic Learning and Intervention Pathways through the Integration of the SDL Module and LLM Agent

Our study highlights the critical role of integrating behavioral intervention frameworks (e.g., BABC) into the conversational logic of LLM agents to support the intervention planning and decision-making process. Practitioners positively recognized the implicit embedding of these frameworks within the LLM agent's dialogue, emphasizing its practical value in guiding behavior management strategies.

Findings from Sections 5.3.1 and 5.3.2revealed distinct yet complementary usage patterns of the Self-Directed Learning (SDL) module and the LLM agent. Specifically, practitioners engaged with the SDL module through a structured learning process: browse/recall → identify a problem → learn → adapt (→ implement). This systematic, textbook-like learning approach was well-suited for non-urgent scenarios where practitioners had time to explore content based on their interests or a child's specific needs. The SDL module thus served as a resource for exploratory and preparatory learning. Additionally, Glitter's context-driven environment encouraged practitioners to engage in problem-solving as a learning opportunity (Fischer & Scharff, 1998). In contrast, interactions with the LLM agent followed a more dynamic and responsive pattern: (encounter a problem →) query → learn → implement (Shalong et al., 2024). This approach enabled practitioners to access immediate, personalized guidance, particularly when managing emergent behavioral challenges or when urgent decisions needed to be made. The LLM agent offered a potential solution to the traditional lack of real-time support in SDL (Roe & Perkins, 2024; Wu, Zhang, Ma, Yue, & Dong, 2024), providing flexible, on-demand assistance. Importantly, practitioners recognized the SDL module and the LLM agent as complementary tools. The SDL module was primarily used for structured learning, quick referencing, and knowledge consolidation, while the LLM agent supported real-time problem-solving and adaptive learning. Their combined use across synchronous and asynchronous intervention scenarios demonstrated the advantages of integrating structured self-learning with adaptive, query-based learning.

A key insight from these findings is the value of designing **dynamic learning pathways** that seamlessly integrate both the SDL module and LLM agent interactions. In the context of the Glitter system, this integrated approach empowered practitioners with greater autonomy in learning (Candy, 1991) and supported self-directed learning as a personalized, goal-driven developmental process (Morris, 2019). The SDL module provided structured learning through virtual case studies (Benedict, Schonder, & McGee, 2013), enabling practitioners to navigate clear learning paths and avoid the frustration that may arise from unstructured content. It also incorporated self-assessment tools, such as quizzes and decision-based outcomes, enabling practitioners to monitor their progress (Knowles, 1975) and recognize knowledge or skill gaps (Benedict et al., 2013). Meanwhile, the LLM agent addressed the challenge of continuously evolving knowledge in behavioral interventions (Garrison, 1997). Its strong information retrieval capabilities, combined with preset behavioral frameworks and user-specific needs, allowed for the dynamic generation of personalized learning tasks (Gotavade, 2024). This adaptability expanded the SDL module's content coverage and met the growing demand for flexible, technology-enhanced learning solutions that support continuous professional development.

By integrating the structured learning approach of the SDL module with the adaptive, real-time support of the LLM agent, the Glitter system fosters a responsive, engaging, and supportive learning environment. Such integration potentially supports practitioners' comprehensive dynamic intervention needs and may also assist them with long-term professional growth.

6.2. Prompting LLM Agents for Better User Experiences in Intervention Contexts

Constructing Users' Role-oriented & Intervention Process-directed Dialogue Logic and Language Style.

While LLMs have been shown to support self-directed learning (Roe & Perkins, 2024), the challenge remains to better meet the individual needs of different practitioners (Tate, Wright, Scott, & Robinson, 2024). The findings from Sections 5.3.2, 5.3.4, and 5.3.5 revealed that the dialogue process, query objectives, and expectations for intervention-related responses varied based on users' roles, role-specific goals, and prior intervention steps (aligned with the BABC stages). This highlighted the importance and approaches of designing tailored LLM agent prompts. First, LLM agents should be utilized to support diverse learning approaches, including knowledge construction, inquiry-based learning, self-assessment, and peer teaching, all of which are reflected in the use case scenarios. Prompt engineering tailored to specific learning contexts is expected to enhance the model's ability to provide effective support compared to general-purpose interactions (Ali, Choy, Divaharan, Tay, & Chen, 2023). Second, role-specific prompts are needed, especially in the context of potentially expanding user groups to other stakeholders. For instance, practitioners often use the agent for lesson preparation or comparing strategies, benefiting from the full query-learn-implement process across BABC stages. In contrast, caregivers likely seek direct support and may not follow the full intervention process. Therefore, the LLM agents could break down intricate behavioral problems into manageable components of structural dialogue logic based on users' roles (i.e., professional vs. non-professional, experienced vs. novice), purposes (inquiry for immediate solutions, compare for better strategies, or prepare for materials), and intervention stages, rather than attempting to complete a full intervention process. In short, the agent's simpler and more accessible language customized to the users' roles could enhance users' comprehension and reduce their cognitive load. Furthermore, participants' expectations regarding the design of an agent's appearance, gestures, and voice align with its anticipated role in fostering a close and engaging communicative environment (Ali et al., 2023). Our field findings indicate that, in this practitioner-focused deployment, informational content and voice were generally prioritized over visual and affective cues for intervention-related tasks. At the same time, affective interaction remains a valuable design dimension, and future work should explore additional technical approaches to better understand its role across different population and task contexts.

Provide Tailored Prompting Question Examples in Tutorials. Another key insight that emerged from the study is the need to offer practitioners and caregivers tutorials or structured support for preparing effective prompts when interacting with the LLM agent. While the system's adaptive dialogue logic allowed for a flexible and open-ended conversation, as suggested in Section 5.3.5, users often struggled to frame their queries in ways that elicited the most useful responses from the agent or to engage in in-depth multi-round conversations. Although we have not evaluated this system with caregivers, this challenge could also occur in the case of non-professional caregivers, who lack the domain-specific knowledge that is necessary to formulate precise or relevant questions. By providing tutorial resources or tailored prompt guidance, the system could help users better understand how to ask questions and interact with the agent based on their intervention goals, current challenges, and the stage they are in within an intervention process, ensuring that they receive the most relevant and useful information. For instance, practitioners might benefit from advanced tutorials that emphasize strategic questioning for refining intervention plans or comparing techniques in multi-round conversations, while caregivers could receive simplified examples tailored to everyday situations they encounter with their children.

LLM Agents in Autistic vs. General Clinical and Instructional Contexts Recent work in education and mental health underscores the growing role of LLMs as instructional and therapeutic partners, highlighting opportunities and risks in both general and inclusive/autistic educational contexts. For instance, studies show that LLMs can act as scaffolding tools that support teaching activities in conceptual understanding, meta-cognitive reflection, and decision-making in complex domains (Kasneci et al., 2023). Emerging research in the inclusive therapeutic area also demonstrates how generative AI can assist with emotional well-being (Ferrara, 2022), behavioral health management (E. C. Stade et al., 2024), and psychotherapies (E. Stade et al., 2023), among other applications. At the same time, researchers caution that the use of LLMs in education raises concerns related to accuracy, pedagogical appropriateness, and the risk of students over-relying on automated feedback (Rudolph, Tan, & Tan, 2023a, 2023b), which are similar issues that occur in inclusive instructional settings. Additionally, an LLM-based supporting tool for stakeholders of autistic children requires consideration of factors beyond general instructional demands, such as supporting highly individualized teaching plans and learning needs, minimizing potential behavioral misunderstandings, and ensuring that AI-generated guidance does not inadvertently reinforce harmful practices. Our findings extend this growing evidence base by illustrating practitioners' desire and demands for adaptive, context-aware guidance that aligns with established behavioral frameworks—an area not yet well addressed in current LLM applications. Unlike prior systems designed primarily for "patient-facing" students or therapeutic clients, our work explicitly centers practitioners' intervention workflows and real-time support, which points to a promising and critical design opportunity within LLM-based educational technologies.

6.3. Mitigating Risks for Practitioners of Autistic Children in LLM-Generated Behavioral Intervention Content

Findings in Section 5.3.4 indicated that LLM agents hold promise for non-professional caregivers, but practitioners also thought they could pose risks, especially for users who may lack formal training in behavioral interventions (i.e., parents). Since practitioners are already well-equipped with professional knowledge, they can effectively draw inspiration from the generated content, which expands their knowledge base or "toolbox," and assess the accuracy of the information provided. However, non-professional caregivers cannot freely or flexibly adapt and implement interventions directly into their children's intervention scenarios. Furthermore, non-professional users may misunderstand intervention content, be unable to recognize poor suggestions, or misapply behavioral strategies.

Beyond these potential concerns, broader ethical and safety considerations in AI-mediated ASD support require explicit attention. In our study, we intentionally implemented strategies, *i.e.,* the BABC framework as the conversation flow with a pre-prompted database and handling for unknown questions (see Section 3.3.2) for safeguard purposes. While preparing the pre-prompted database, we also removed sensitive data during processing, and no personal data was collected. Although no practitioners identified mistakes or reported harmful content in the generated answers in our study (consistent with prior research (McFayden et al., 2024)), recent research cautions that LLM systems can introduce misinformation, over-confidence, or misinterpretation risks in autistic interventions, therapeutic and educational contexts (Adako, Adeusi, & Alaba, 2025; Jaliaawala & Khan, 2020; G. Li, Zarei, Alibakhshi, & Labbafi, 2024). Similarly, broader reviews of AI in ASD assessment and intervention highlight risks related to model opacity, inappropriate generalization, and unintended behavior when AI tools are applied without human oversight (Abualhoul, 2025; Kotsi, Handrinou, Iatraki, & Soulis, 2025). To support safe deployment in clinical or school environments, e.g., the autistic training scenario in this work, future systems could consider incorporating layered risk-mitigation strategies, such as automated detection of potentially unsafe outputs, built-in uncertainty or warning messaging for non-professional caregivers, and privacy-preserving mechanisms (e.g., data minimization, on-device processing, and transparent data-use guidelines). Implementing these safeguards can potentially enhance the reliability and ethical integrity of LLM-based tools as they transition into real-world behavioral intervention settings.

One other recurring theme was the importance of privacy and trust for both practitioners and caregivers. Practitioners highly preferred anonymity while using the app, and preferred that no autism-related words be explicitly presented in the conversation or user interfaces to maintain discretion. Such anonymity made them feel comfortable and secure in sharing autistic children's private situations when querying individual intervention information. While practitioners generally trusted the agent due to Glitter's "dedicated content and professional appearance", prior work suggested that ensuring privacy features are explicitly communicated would reinforce clear and transparent dialogues with virtual agents. Therefore, adding privacy guidelines for data collection, utilization, and generation in the current autistic intervention context could be considered and implemented in future work.

6.4. Design Implications for Expanding LLM Support Across Educational Stakeholders

Our findings highlight promising directions for future work by underscoring the broader ecosystem of stakeholders who could benefit from an LLM-based behavioral intervention agent. Practitioners emphasized that such tools extend beyond specialist use, offering value to teachers in special education and general K–12 classrooms who often juggle large class sizes, diverse student needs, and limited time to revisit behavioral intervention techniques. These insights suggest opportunities for designing agents that support rapid knowledge refreshment and classroom-embedded decision-making. Participants also noted the potential impact on teachers in low-socioeconomic or remote regions, who typically lack access to training and peer support, highlighting the need for equitable and scalable deployment models. Additionally, reflections from P6 and P10 illustrated how novice practitioners, such as university students transitioning into professional roles, could benefit from scaffolded, real-world examples that bridge the gap between theoretical coursework and practical intervention work. Future research could further explore context-adaptive agent features, stakeholder-specific interface designs, and mechanisms that support novices and educators, ultimately broadening the reach and inclusivity of LLM-supported behavioral intervention training.

Although our findings suggest that the agent could be a helpful assistant to non-professional users, the exclusion of these users limits the generalizability of the findings to other user groups, who may use it differently and encounter different challenges when using the LLM agent. Additionally, the current study did not include parents and other caregivers (e.g., grandparents, social workers) as participants, which is an important consideration for future work, given the potential negative consequences that generative responses may cause in misleading non-professional users.

6.5. Limitations and Future Work

This study has several limitations that should be considered when interpreting the findings. First, the practitioner sample was relatively small (N=11) and demographically homogeneous, characterized by a strong gender imbalance (10/11 female) and a narrow age range (20–29 years). While this composition reflects the current workforce structure of China's rapidly expanding special education sector (Hu, Liu, & Zhai, 2023; Ministry of Education of the People's Republic of China, 2022; Wang, Dong, & Cao, 2022), it limits the generalizability of our findings. Future studies could adopt multi-site recruitment across diverse regions and special education institutions (e.g., public schools, private intervention centers, and community-based organizations) to examine whether the findings generalize across different practice contexts.

Second, the one-week deployment period was relatively short and may have introduced novelty effects. Long-term, longitudinal studies are necessary to investigate the sustained use and integration of this approach into everyday practice. Additionally, as an exploratory observational study, this work did not include a control group. Future research aimed at evaluating intervention effectiveness more rigorously should incorporate comparison groups (e.g., traditional materials or alternative tools) to strengthen causal inference.

Third, there was an absence of objective, performance-based indicators (Kazdin, 2017). Although we conducted an expert-based assessment of response quality, accuracy, and actionability, this assessment focused on content-level properties rather than practitioners' performance or intervention outcomes. At the practitioner level, this exploratory, in-the-wild study relied primarily on self-report questionnaires, interviews, and usage logs, as our goal was to understand practitioners' use cases, interaction patterns, and perceptions of an LLM-based decision-support agent in real-world settings. Accordingly, we did not assess changes in practitioners' decision-making performance. Future work could incorporate objective practitioner-level measures, such as expert evaluation of intervention plans or task-based assessments of decision accuracy and efficiency. At the child level, we did not assess downstream behavioral intervention outcomes. Child-level change in autism interventions depends on multiple interacting factors beyond decision support, including implementation fidelity, contextual constraints, and cumulative effects over time. Nevertheless, we consider child behavioral outcomes to be an important ultimate indicator of long-term effectiveness. Future work should therefore integrate child-level behavioral measures with practitioner-centered performance indicators to enable a more comprehensive evaluation.

Fourth, we did not conduct systematic safety and robustness testing of the LLM-generated content. Although no practitioners reported harmful or inappropriate suggestions during the study, this small-scale deployment is insufficient to rule out risks such as misinformation, overconfident advice, or inappropriate generalization in sensitive behavioral intervention contexts. Future work should incorporate structured safety evaluations, including adversarial prompting, boundary-case testing, and systematic analysis of failure modes. In addition, mechanisms such as uncertainty communication, built-in disclaimers, and human-in-the-loop review should be explored to mitigate potential risks when deploying practitioner-facing LLM systems in real-world settings.

Additionally, the LLM agent relied on a small sample intervention dataset and prompt set for generating responses, which may not fully capture the diversity of real-world behavioral issues. The inclusion of only the ABA framework and a few intervention techniques (i.e., BABC, DTT) further restricts the scope of the generated intervention strategies and the potential interactions and dynamics between users and the agent. Therefore, future work could also consider expanding the dataset and iterating the prompts with more comprehensive intervention techniques and frameworks. Caregiver groups are also needed to evaluate the LLM agent's applicability across a broader range of contexts and, in the long term, cultures and world populations.

7. Conclusion

In conclusion, this study provides an exploratory investigation into the role of LLM-based agents like Glitter in supporting practitioners and (potentially) caregivers in behavioral interventions for autistic children from an iterative design process and a one-week field study. Our findings provide field-informed insights into the design of more accessible LLM-powered virtual agents for both practitioners and caregivers by exploring dynamic learning pathways, tailored LLM dialogue logic, the influences of multi-modality features, and potential opportunities and challenges.

8. Acknowledgment

This research is funded in part by the Guangdong Science and Technology Plan (2023A1111120012) and Duke Kunshan University Synear and Wang-Cai Seed Grant (21KCNGO016). We thank the Research Fund for International Scientists (Project No. W2533160), National Natural Science Foundation of China (NSFC) for supporting this work. We also acknowledge the support from Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007). This work is also supported by the "111 Center" (Project No. D25008).

9. Disclosure Statement

Xin Tong, Ming Li, Liwen He, Weibo Li, Yutong Ren, Zhaowen Deng, and Ziheng Tang hold a granted patent titled "Multimodal Emotional Interaction System and Method" (Patent No. ZL 2024 1 0902796.3, CN 118888086 B), assigned to Duke Kunshan University. This may be considered a potential competing interest. All other authors declare that they have no known financial or non-financial competing interests that could have appeared to influence the work reported in this paper.

References

Abualhoul, M. (2025). Use artificial intelligence in autism spectrum disorder: a comprehensive review in assessment, intervention, and data collection. *Journal of Information Technology, Cybersecurity, and Artificial Intelligence*, *2*(2), 20–25. doi:

Adako, O. P., Adeusi, O. C., & Alaba, P. A. (2025). Ai in autism education: a review of collaborative and longitudinal approaches. *Disability and Rehabilitation: Assistive Technology*, 1–26. Retrieved from `https://doi.org/10.1080/17483107.2025.2579825` doi:

Alexander, J. L., Ayres, K. M., & Smith, K. A. (2015). Training teachers in evidence-based practice for individuals with autism spectrum disorder: A review of the literature. *Teacher Education and Special Education*, *38*(1), 13–27. doi:

Ali, F., Choy, D., Divaharan, S., Tay, H. Y., & Chen, W. (2023). Supporting self-directed learning and self-assessment using teachergaia, a generative ai chatbot application: Learning approaches and prompt engineering. *Learning: Research and Practice*, *9*(2), 135–147. doi:

Alkinj, I., Pereira, A., & Santos, P. (2022, 05). The effects of an educational program based on modeling and social stories on improvements in the social skills of students with autism. *Heliyon*, *8*(5), e09289. doi:

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: Dsm-5* (Vol. 5). Washington, DC: American Psychiatric Association. doi:

Babalola, T., Sanguedolce, G., Dipper, L., & Botting, N. (2024). Barriers and facilitators of healthcare access for autistic children in the uk: A systematic review. *Review Journal of Autism and Developmental Disorders*. doi:

Bak, M. Y. S., Plavnick, J. B., Dueñas, A. D., Brodhead, M. T., Avendaño, S. M., Wawrzonek, A. J., ... Oteto, N. (2021). The use of automated data collection in applied behavior analytic research: A systematic review. *Behavior Analysis: Research and Practice*, *21*(4), 376–405. doi:

Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). New York: Wiley. doi:

Benedict, N., Schonder, K., & McGee, J. (2013). Promotion of self-directed learning using virtual patient cases. *American Journal of Pharmaceutical Education*, *77*(7), 151. doi:

Blair, B. J., & Shawler, L. A. (2020). Developing and implementing emergent responding training systems with available and low-cost computer-based learning tools: Some best practices and a tutorial. *Behavior Analysis in Practice*, *13*(2), 509–520. doi:

Bonis, S. (2016). Stress and parents of children with autism: A review of literature. *Issues in Mental Health Nursing*, *37*(3), 153–163. doi:

Bradshaw, J., Wolfe, K., Hock, R., & Scopano, L. (2022). Advances in supporting parents in interventions for autism spectrum disorder. *Pediatric Clinics of North America*, *69*(4), 645–656. doi:

Bradshaw, P., Pickett, C., van Driel, M. L., Brooker, K., & Urbanowicz, A. (2021). 'autistic' or 'with autism'?: Why the way general practitioners view and talk about autism matters. *Australian Journal of General Practice*, *50*(3), 104–108. (Publisher's DOI may not resolve in some systems. Article available at: https://www1.racgp.org.au/ajgp/2021/march/autistic-or-with-autism) doi:

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. doi:

Bromley, A. R. (2023). Flexibility within fidelity: A narrative review of practitioner modifications to child welfare interventions. *Children and Youth Services Review*, *149*, 106908. doi:

Brooke, J. (1996). Sus: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). CRC Press. doi:

Buchanan, S. M., & Weiss, M. J. (2006). *Applied behaviour analysis and autism: An introduction.* Autism New Jersey. (No DOI available for this manual.)

Caldwell, T. D., Schreck, K. A., Spisak, A. N., & Katz, J. I. (2025). Barriers to use of experimental analysis in applied behavior analysis clinical practice. *Behavior Analysis in Practice*, *18*(3), 667–680. doi:

Candy, P. C. (1991). *Self-direction for lifelong learning: A comprehensive guide to theory and practice.* San Francisco: Jossey-Bass.

Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., ... Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions*, *4*(1), 4–16. doi:

Carr, J. E., & Firth, A. M. (2005). The verbal behavior approach to early and intensive behavioral intervention for autism: A call for additional empirical support. *Journal of Early and Intensive Behavior Intervention*, *2*(1), 18. doi:

Charlop-Christy, M. H., Carpenter, M., Le, L., LeBlanc, L. A., & Kellet, K. (2002). Using the picture exchange communication system (pecs) with children with autism: Assessment of pecs acquisition, speech, social-communicative behavior, and problem behavior. *Journal of Applied Behavior Analysis*, *35*(3), 213–231. doi:

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 213–218). doi:

Chown, N., Shaw, S., Doherty, M., Johnson, M., Krupa, J., Martin, N., & Brooker-Corcoran, M. (2023). General practitioner autism training and mandatory medical training: A cross-sectional study of gps' knowledge, attitudes and practices. *Canadian Journal of Educational and Social Studies*, *3*(1), 1–16. (No DOI assigned by this journal.)

Clarke, L., & Fung, L. K. (2022). The impact of autism-related training programs on physician knowledge, self-efficacy, and practice behavior: A systematic review. *Autism*, *26*(7), 1626–1640. doi:

Curran, C., Roberts, R., Gannoni, A., & Jeyaseelan, D. (2024). Training and educational pathways for clinicians (post-graduation) for the assessment and diagnosis of autism spectrum disorders: A scoping review. *Journal of Autism and Developmental Disorders*, 1–21. doi:

Deng, C., Lai, S., Zhou, C., Bao, M., Yan, J., Li, H., ... Wang, Y. (2024). Asd-chat: An innovative dialogue intervention system for children with autism based on llm and vb-mapp. *arXiv preprint arXiv:2409.01867*. doi:

Dominick, K. C., Davis, N. O., Lainhart, J., Tager-Flusberg, H., & Folstein, S. (2007). Atypical behaviors in children with autism and children with a history of language impairment. *Research in Developmental Disabilities*, *28*(2), 145–162. doi:

Eckes, T., Buhlmann, U., Holling, H.-D., & Möllmann, A. (2023). Comprehensive aba-based interventions in the treatment of children with autism spectrum disorder–a meta-analysis. *BMC Psychiatry*, *23*(1), 133. doi:

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009a). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology*, *38*(3), 439–450. doi:

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009b). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology*, *38*(3), 439–450. doi:

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, *5*(1), 80–92. doi:

Ferrara, A. (2022). *Empowering emotional well-being through a llm-based chatbot: a comparative study with the standard journaling technique* (Unpublished master's thesis). Politecnico di Torino. (Master's Thesis. No DOI assigned.)

Fischer, G., & Scharff, E. (1998). Learning technologies in support of self-directed learning. *Journal of Interactive Media in Education*, *1998*(2), 4. doi:

Florencio, D., & Herley, C. (2007). A large-scale study of web password habits. In *Proceedings of the 16th international conference on world wide web* (pp. 657–666). ACM. doi:

Fox, L., Vaughn, B. J., Wyatte, M. L., & Dunlap, G. (2002). "we can't expect other people to understand": Family perspectives on problem behavior. *Exceptional Children*, *68*(4), 437–450. doi:

Garrison, D. R. (1997). Self-directed learning: Toward a comprehensive model. *Adult Education Quarterly*, *48*(1), 18–33. doi:

Glenn, E., Taiwo, A., Arbuckle, S., Riehl, H., & McIntyre, L. L. (2023). Self-directed web-based parent-mediated interventions for autistic children: A systematic review. *Review Journal of Autism and Developmental Disorders*, *10*(4), 505–522. doi:
Gotavade, T. S. (2024). Artificial intelligence ecosystem for automating self-directed teaching. *arXiv preprint arXiv:2411.07300*. doi:
Granpeesheh, D., Tarbox, J., & Dixon, D. R. (2009). Applied behavior analytic interventions for children with autism: A description and review of treatment research. *Annals of Clinical Psychiatry*, *21*(3), 162–173. (No DOI available. PMID: 19723358)
Hamad, C. D., Serna, R. W., Morrison, L., & Fleming, R. (2010). Extending the reach of early intervention training for practitioners: A preliminary investigation of an online curriculum for teaching behavioral intervention knowledge in autism to families and service providers. *Infants & Young Children*, *23*(3), 195–208. doi:
Hammond, A. W., Morris, J. R., Gabrielsen, T., Smith, T. B., & Medsker, N. (2025). A systematic review of digital activity schedule use in individuals with autism spectrum disorder and intellectual disability. *Journal of Intellectual & Developmental Disability*, 1–15. doi:
Hu, Y., Liu, Y., & Zhai, D. (2023). The development of discipline teaching methodology faculty in chinese higher education institutions: Current status, challenges, and countermeasures []. *Journal of Educational Studies []*, *19*(5), 109–119. doi:
Ibañez, L. V., Kobak, K., Swanson, A., Wallace, L., Warren, Z., & Stone, W. L. (2018). Enhancing interactions during daily routines: A randomized controlled trial of a web-based tutorial for parents of young children with asd. *Autism Research*, *11*(4), 667–678. doi:
Jaliaawala, M. S., & Khan, R. A. (2020). Can autism be catered with artificial intelligence-assisted intervention technology? a comprehensive survey. *Artificial Intelligence Review*, *53*(2), 1039–1069. Retrieved from `https://doi.org/10.1007/s10462-019-09686-8` doi:
Jang, J., Moharana, S., Carrington, P., & Begel, A. (2024). "it's the only thing i can trust": Envisioning large language model use by autistic workers for communication assistance. In *Proceedings of the chi conference on human factors in computing systems* (pp. 1–18). ACM. doi:
Kapp, S. K., Gillespie-Lynch, K., Sherman, L. E., & Hutman, T. (2013). Deficit, difference, or both? autism and neurodiversity. *Developmental Psychology*, *49*(1), 59–71. doi:
Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... others (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. doi:
Kazdin, A. E. (2017). Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. *Behaviour Research and Therapy*, *88*, 7–18. doi:
Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? perspectives from the uk autism community. *Autism*, *20*(4), 442–462. doi:
Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers.* New York: The Adult Education Company.
Kotsi, S., Handrinou, S., Iatraki, G., & Soulis, S.-G. (2025). A review of artificial intelligence interventions for students with autism spectrum disorder. *Disabilities*, *5*(1), 7. doi:
Lazar, J., Feng, J. H., & Hochheiser, H. (2017). *Research methods in human-computer interaction.* Morgan Kaufmann. doi:
Lee, I.-J., & Hsu, H.-T. (2023). Applied the augmented reality technology combined with social stories strategies and computational thinking games to improve the social skills of children with asd. *Interactive Learning Environments*. doi:
Lee, L.-C., Harrington, R. A., Louie, B. B., & Newschaffer, C. J. (2008). Children with autism: Quality of life and parental concerns. *Journal of Autism and Developmental Disorders*, *38*(5), 1147–1160. doi:
Li, G., Zarei, M. A., Alibakhshi, G., & Labbafi, A. (2024). Teachers and educators' experiences and perceptions of artificial-powered interventions for autism groups. *BMC Psychology*, *12*(1), 199. Retrieved from `https://doi.org/10.1186/s40359-024-01664-2` doi:
Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and

meta-analysis of ai-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine*, *6*(1), 236. doi:

Li, Z., Babar, P. P., Barry, M., & Peiris, R. L. (2024). Exploring the use of large language model-driven chatbots in virtual reality to train autistic individuals in job communication skills. In *Extended abstracts of the chi conference on human factors in computing systems* (pp. 1–7). ACM. doi:

Lim, J. M. (2020). Emotion regulation and intervention in adults with autism spectrum disorder: A synthesis of the literature. *Advances in Autism*, *6*(1), 48–62. doi:

Lindgren, N. (2025). *An evaluation of using linked digital activity schedules to teach socio-dramatic play* (Doctor of Philosophy (PhD), Utah State University). (Committee Chair: Thomas S. Higbee) doi:

Lovaas, O. I., Schreibman, L., & Koegel, R. L. (1974). A behavior modification approach to the treatment of autistic children. *Journal of Autism and Childhood Schizophrenia*, *4*(2), 111–129. doi:

Ma, Z., Mei, Y., & Su, Z. (2023). Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. In *Amia annual symposium proceedings* (Vol. 2023, pp. 1105–1114). (No DOI assigned for this proceedings volume.)

Mahon, D., Dunworth, C., McSharry, J., Holloway, J., & Lydon, H. (2025). Barriers and facilitators to staff's implementation of behaviour support plans in community settings. *Journal of Applied Research in Intellectual Disabilities*, *38*(2). doi:

Makrygianni, M. K., Gena, A., Katoudi, S., & Galanis, P. (2018). The effectiveness of applied behavior analytic interventions for children with autism spectrum disorder: A meta-analytic study. *Research in Autism Spectrum Disorders*, *51*, 18–31. doi:

Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with autistic spectrum disorders. *Research in Autism Spectrum Disorders*, *4*(4), 577–593. doi:

Matson, J. (2009). Aggression and tantrums in children with autism: A review of behavioral treatments and maintaining variables. *Journal of Mental Health Research in Intellectual Disabilities*, *2*(3), 169–187. doi:

Matson, J. L., & Nebel-Schwalm, M. (2007). Assessing challenging behaviors in children with autism spectrum disorders: A review. *Research in Developmental Disabilities*, *28*(6), 567–579. doi:

Mattson, S. L., Higbee, T. S., Aguilar, J., Nichols, B., Campbell, V. E., Nix, L. D., ... Lewis, K. (2020). Creating and sharing digital aba instructional activities: A practical tutorial. *Behavior Analysis in Practice*, *13*(4), 772–798. doi:

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–23. doi:

McFayden, T. C., Bristol, S., Putnam, O., & Harrop, C. (2024). Chatgpt: Artificial intelligence as a potential tool for parents seeking information about autism. *Cyberpsychology, Behavior, and Social Networking*, *27*(2), 135–148. doi:

McGreevey III, J. D., Hanson III, W., & Koppel, R. (2020). Clinical, legal, and ethical aspects of artificial intelligence–assisted conversational agents in health care. *JAMA*, *324*(6), 552–553. doi:

McIntyre, L. L., & Phaneuf, L. K. (2008). A three-tier model of parent education in early childhood: Applying a problem-solving model. *Topics in Early Childhood Special Education*, *27*(4), 214–222. doi:

Ministry of Education of the People's Republic of China. (2022, January). *The "14th five-year" special education promotion and improvement plan (2022-2025).* GOV.CN.

Mishra, R., & Welch, K. C. (2024). Towards scalable robotic intervention of children with autism spectrum disorder using llms. *arXiv preprint arXiv:2402.00260*. doi:

Morrier, M. J., Hess, K. L., & Heflin, L. J. (2011). Teacher training for implementation of teaching strategies for students with autism spectrum disorders. *Teacher Education and*

Special Education, *34*(2), 119–132. doi:
Morris, T. H. (2019). Self-directed learning: A fundamental competence in a rapidly changing world. *International Review of Education*, *65*(4), 633–653. doi:
National Research Council. (2001). *Educating children with autism.* Washington, DC: National Academies Press. doi:
O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. doi:
Oh, K. J., Lee, D., Ko, B., & Choi, H. J. (2017, 5). A chatbot for psychiatric counseling in mental healthcare service based on emotional dialogue analysis and sentence generation. In *2017 18th ieee international conference on mobile data management (mdm)* (pp. 371–376). IEEE. doi:
Pan, Q., & Meng, Z. (2024, 2). Hybrid uncertainty calibration for multimodal sentiment analysis. *Electronics*, *13*(3), 662. doi:
Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive aba-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 60–69. doi:
Pettersson, K., Millroth, P., Giannotta, F., Liedgren, P., Lyon, A. R., Hasson, H., & von Thiele Schwarz, U. (2025). Outcome preferences in fidelity-adaptation scenarios across evidence-based parenting programs: A discrete choice experiment. *Implementation Science*, *20*(1), 10. doi:
Pickard, K., Rowless, S., & Ingersoll, B. (2019). Understanding the impact of adaptations to a parent-mediated intervention on parents' ratings of perceived barriers, program attributes, and intent to use. *Autism*, *23*(2), 338–349. doi:
Pomales-Ramos, A., Tokish, H., Howard, M., Straiton, D., & Ingersoll, B. (2023). A mixed-methods examination of clinicians' perceived barriers to telehealth delivered applied behavior analysis. *Frontiers in Psychology*, *14*, 1173644. doi:
Qiu, H., & Lan, Z. (2024). *Interactive agents: Simulating counselor-client psychological counseling via role-playing llm-to-llm interactions.* doi:
Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (eibi) for young children with autism spectrum disorders (asd). *Cochrane Database of Systematic Reviews*(10). doi:
Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2014). Early intensive behavioral intervention (eibi) for young children with autism spectrum disorders (asd): A systematic review. *Campbell Systematic Reviews*, *10*(1), 1–116. doi:
Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the ucla young autism project model. *Journal of Autism and Developmental Disorders*, *39*, 23–41. doi:
Ren, X., Bai, Y., Duan, H., Fan, L., Fei, E., Wu, G., ... Zhai, G. (2024a). Chatasd: Llm-based ai therapist for asd. In *Digital multimedia communications (iftc 2023)* (pp. 312–324). Springer. doi:
Ren, X., Bai, Y., Duan, H., Fan, L., Fei, E., Wu, G., ... Zhai, G. (2024b). Chatasd: Llm-based ai therapist for asd. In G. Zhai, J. Zhou, L. Ye, H. Yang, P. An, & X. Yang (Eds.), *Digital multimedia communications* (pp. 312–324). Singapore: Springer Nature Singapore. doi:
Rivard, M., Mello, C., Sanchez, C., Grenier-Martin, J., Lefebvre, C., Forget, J., ... Morin, D. (2025). Barriers and facilitators to implementing prevent-teach-reinforce for young children in community-based early intervention services for autism. *Evaluation and Program Planning*, *112*, 102657. doi:
Roe, J., & Perkins, M. (2024). Generative ai in self-directed learning: A scoping review. *arXiv preprint arXiv:2411.07677*. doi:
Rudolph, J., Tan, S., & Tan, S. (2023a). Chatgpt: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning & teaching*, *6*(1), 342–363. doi:
Rudolph, J., Tan, S., & Tan, S. (2023b). War of the chatbots: Bard, bing chat, chatgpt, ernie

and beyond. the new ai gold rush and its impact on higher education. *Journal of Applied Learning & Teaching*, *6*(1), 364–389. doi:

Sam, A. M., Cox, A. W., Savage, M. N., Waters, V., & Odom, S. L. (2020). Disseminating information on evidence-based practices for children and youth with autism spectrum disorder: Afirm. *Journal of Autism and Developmental Disorders*, *50*, 1931–1940. doi:

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., … others (2020). Project aim: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin*, *146*(1), 1–29. doi:

Shalong, W., Yi, Z., Bin, Z., Ganglei, L., Jinyu, Z., Yanwen, Z., … Feng, R. (2024). Enhancing self-directed learning with custom gpt ai facilitation among medical students: A randomized controlled trial. *Medical Teacher*, 1–8. doi:

Skinner, B. F. (1971). Operant conditioning. *The Encyclopedia of Education*, *7*, 29–33. (No DOI available for this encyclopedia entry.)

Slanzi, C. M., & Fernand, J. K. (2024). On the use and benefits of electronic data collection systems: A tutorial on countee. *Behavior Analysis in Practice*, *17*(4), 1228–1237. doi:

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities*, *16*(2), 86–92. doi:

Spreckley, M., & Boyd, R. (2009). Efficacy of applied behavioral intervention in preschool children with autism for improving cognitive, language, and adaptive behavior: A systematic review and meta-analysis. *The Journal of Pediatrics*, *154*(3), 338–344. doi:

Stade, E., Stirman, S. W., Ungar, L. H., Schwartz, H. A., Yaden, D. B., Sedoc, J., … others (2023). Artificial intelligence will change the future of psychotherapy: A proposal for responsible, psychologist-led development. *PsyArXiv*. doi:

Stade, E. C., Stirman, S. W., Ungar, L. H., Boland, C. L., Schwartz, H. A., Yaden, D. B., … Eichstaedt, J. C. (2024). Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *NPJ Mental Health Research*, *3*(1), 12. doi:

Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, *20*(2), 66–79. doi:

Sutherland, R., Trembath, D., & Roberts, J. (2018). Telehealth and autism: A systematic search and review of the literature. *International Journal of Speech-Language Pathology*, *20*(3), 324–336. doi:

Tang, Y., Chen, L., Chen, Z., Chen, W., Cai, Y., Du, Y., … Sun, L. (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* (pp. 1–20). ACM. doi:

Tarver, J., Palmer, M., Webb, S., Scott, S., Slonims, V., Simonoff, E., & Charman, T. (2019). Child and parent outcomes following parent interventions for child emotional and behavioral problems in autism spectrum disorders: A systematic review and meta-analysis. *Autism*, *23*(7), 1630–1644. doi:

Tate, B., Wright, J., Scott, E., & Robinson, S. (2024). Dynamic parameter morphogenesis for adaptable task specialization in large language models to self-directed learning. (No DOI found. Citation appears unverifiable/hallucinated.)

Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., … 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. doi:

Villamin, G., & Luppicini, R. (2024). Digital interventions using mobile technologies for life skills development of learners with autism spectrum disorder: A scoping review. *International Journal of Developmental Disabilities*. doi:

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical Psychology Review*, *30*(4), 387–399. doi:

Vismara, L. A., & Rogers, S. J. (2010). Behavioral treatments in autism spectrum disorder:

What do we know? *Annual Review of Clinical Psychology*, *6*(1), 447–468. doi:

Wang, Y., Dong, Y., & Cao, J. (2022). Reflections and suggestions on the cultivation of special education professionals in china under the background of professional certification: A text analysis based on training programs from 25 universities []. *Chinese Journal of Special Education []*(5), 68–76. doi:

Wang, Y., Gu, Y., Yin, Y., Han, Y., Zhang, H., Wang, S., … Quan, D. (2023). Multimodal transformer augmented fusion for speech emotion recognition. *Frontiers in Neurorobotics*, *17*, 1181598. doi:

Wilkenfeld, D. A., & McCarthy, A. M. (2020). Ethical concerns with applied behavior analysis for autism spectrum "disorder". *Kennedy Institute of Ethics Journal*, *30*(1), 31–69. doi:

Wood, J. J., Wood, K. S., Rosenau, K. A., Cho, A. C., Johnson, A. R., Muscatello, V. S., … others (2024). Practitioner adherence and competence in meya, a free online self-instruction program in modular psychotherapy and counseling for children's autism-related clinical needs. *Journal of Autism and Developmental Disorders*. doi:

World Health Organization. (2018). *Rehabilitation in health systems: Guide for action.* Geneva: World Health Organization.

Wu, D., Zhang, S., Ma, Z., Yue, X.-G., & Dong, R. K. (2024). Unlocking potential: Key factors shaping undergraduate self-directed learning in ai-enhanced educational environments. *Systems*, *12*(9), 332. doi:

Zaidman-Zait, A., Mirenda, P., Duku, E., Szatmari, P., Georgiades, S., Volden, J., … others (2014). Examination of bidirectional relationships between parent stress and two types of problem behavior in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *44*, 1908–1917. doi:

10. Author Biographies

Xin Tong is an Assistant Professor in the Computational Media and Arts (CMA) and Robotics and Autonomous Systems (ROAS) Thrusts at the Hong Kong University of Science and Technology (Guangzhou). Her research focuses on intelligent human-computer interaction (HCI), human-AI augmentation, and human-robot interaction.

Liwen He is a Ph.D. candidate in Design at Tsinghua University. She received her bachelor's and master's degrees from Beihang University. Her research focuses on human–computer interaction (HCI), human–robot interaction, and affective haptics.

Zhaowen Deng is a Master's student in Computer Science at Dartmouth College. His research interests include artificial intelligence, human–computer interaction, and intelligent agent systems, with a focus on designing interactive AI systems that enhance collaboration between humans and intelligent technologies.

Weibo Li received the M.S. degree from Taiyuan University of Technology. He was a research intern at Duke Kunshan University. His research focuses on multimodal emotion recognition and model acceleration.

Ziheng Tang received his B.E. degree from Harbin Institute of Technology in 2020. He is currently a Ph.D. candidate in the Department of Mechanical Engineering, Tsinghua University. His research interests include software and hardware development for novel MEMS sensors based on low-dimensional materials.

Yixuan Li is a master's student at Tsinghua University. Her research interests include human-computer interaction, tangible interaction, and AI-assisted interactive systems. Her work has been published at conferences such as CHI and UIST.

Yutong (Tony) Ren is a first-year Ph.D. student in Computer Science at Virginia Tech. His research interests are in the areas of integrating AI within AR/VR environments to develop innovative interactions and user interfaces that enhance adaptability, usability, and user experience.

Matthew Louis Mauriello is an Assistant Professor in the Department of Computer and Information Sciences at the University of Delaware. His research focuses on human–computer interaction, human-centered AI, and interactive systems for health, education, and sustainability, with an emphasis on designing intelligent systems that amplify human expertise.

Ming Li received his Ph.D. in Electrical Engineering from the University of Southern California in 2013. He is currently a Professor of Electrical and Computer Engineering at Duke Kunshan University. His research interests are in the areas of audio, speech and language processing as well as multimodal behavior signal processing.