Open Journal of Social Sciences, 2026, 14(6), 31-63
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Applications, Effectiveness, and Implementation Features of AI-Based Interventions in Early Childhood Special Education: A Systematic Review

Hibah Alharbi

Department of Psychology, University of Bisha, Bisha, Saudi Arabia
Email: halharbi@ub.edu.sa

How to cite this paper: Alharbi, H. (2026). Applications, Effectiveness, and Implementation Features of AI-Based Interventions in Early Childhood Special Education: A Systematic Review. *Open Journal of Social Sciences, 14,* 31-63.
https://doi.org/10.4236/jss.2026.146002

Received: April 26, 2026
Accepted: May 30, 2026
Published: June 2, 2026

Abstract

Artificial intelligence (AI) is increasingly used in preschool special education to support communication, social, and cognitive development, though evidence remains fragmented. This review synthesized child-level effects of AI interventions for young children, including preschool-aged and mixed developmental profiles. The author followed PRISMA 2020 guidelines, applied a PICOS framework, and searched six databases (ERIC, PsycINFO, Scopus, Web of Science, PubMed, IEEE Xplore) from January 2010 to January 2026. Quality was appraised using Cochrane Risk of Bias 2 (RoB 2) for randomized controlled trials and Joanna Briggs Institute (JBI) checklists for other designs. Heterogeneity precluded meta-analysis; therefore, narrative and thematic synthesis were used. The initial search yielded 1472 records. After removing 186 duplicates, 1286 records were screened. Following eligibility assessment, 31 studies met inclusion criteria. Most studies focused on autism spectrum disorder. The most consistent effects were improvements in joint attention and imitation during robot-mediated sessions. Social outcomes improved for emotion recognition, perspective-taking, and adaptive functioning, with maintenance demonstrated in several studies. Cognitive outcomes showed positive trends for attention and engagement but less conclusive evidence. Effective implementations commonly used 20 - 30-minute sessions delivered approximately 10 times. Risk of bias was often moderate due to small samples, short durations, and heterogeneous measures. AI supports targeted gains when aligned to specific skills and embedded within structured routines with adult scaffolding. Larger, longer, culturally responsive trials are needed to confirm durability, generalization, and equity of impact. Importantly, the synthesis distinguishes between direct child-level intervention effects and contextual or system-level applications (e.g., screening, feasibility, and validation studies), which do not constitute efficacy evidence.

Keywords

Artificial Intelligence, Autism Spectrum Disorder, Child, Preschool, Developmental Disabilities, Education, Special, Early Intervention, Robots, Augmented Reality, Computer-Assisted Instruction, Attention, Imitative Behavior, Communication

1. Introduction

Artificial intelligence (AI) is increasingly embedded in early childhood education, where it is used to personalize learning, scaffold communication, and support data-informed decision-making in inclusive settings (Dore & Dynia, 2020). Studies on technology and media application in preschool classrooms have reported on the use of digital tools and their potential in helping diverse learners (Walan, 2025). The most developed applications in preschool special education are social assistive robots, which attract joint attention and recognize emotions; vocabulary-based speech recognition systems; and adaptive platforms, which personalize early literacy practice (Su & Yang, 2022; Telisheva et al., 2022). It has been documented that empirical research has shown that preschoolers with autism spectrum disorder have improved engagement and child-level results, including functional play (Eapen et al., 2013; Estes et al., 2015). The clinical outcomes research has shown that the preschool children in community settings can be provided with early intervention programs that can have significant developmental benefits (De Belen et al., 2023). The developments are topical since intervention in the preschool period is linked to the greatest benefits (Estes et al., 2015; Nan, 2020).

Humanoid robotic platforms, like NAO, Pepper, and CommU, constitute an important modality of long-term and home-based applications that can show sustained user involvement and measurable change in the results of social communication. An example is that Scassellati et al. (2018) demonstrated that triadic sessions with an autonomous social robot daily during a period of over thirty days resulted in significant gains in joint attention and increased caregiver-reported communication behaviors. An extended scoping review of a decade of studies of human-NAO interaction also reported the development of robotic interventions in the context of therapeutic, educational, and assistive applications (Amirova et al., 2021). Furthermore, Kumazaki et al. (2018) held that there were strong carryover effects in which enhanced joint attention by children using the CommU robot effectively transferred to the next human interaction. These studies indicate that properly designed robotic interfaces (Tamaral et al., 2025) support scalable assessment and classroom integration, particularly through automated engagement metrics that dynamically adjust robot behavior to meet individual children's needs effectively (Rudovic et al., 2017; Santos et al., 2021).

Educational frameworks within this field are rapidly expanding to include automated engagement metrics and robotic coaches designed for motor, social, and cognitive training in children with autism spectrum disorder (Rudovic et al., 2017; Santos et al., 2021). Early detection and progress monitoring are provided through special procedures such as eye-tracking studies (De Belen et al., 2023) and robot-mediated screening systems such as Q-CHAT-NAO ((Romero-García et al., 2021). Moreover, interventions in schools with toy robots (Ghiglino et al., 2021) and earlier studies in bilingual human-AI reading settings identify the language learning opportunities of various preschoolers (Feng & Wang, 2023). These heterogeneous technological strands represent an omnifarious AI toolkit for early childhood special education, which is based on assessment, intervention, and program support. When these tools are incorporated into the natural classroom setting, teachers have an opportunity to use high-frequency data to fine-tune the instructional strategy. The synthesis of these applications points toward a future where AI and human instruction work in tandem to optimize developmental trajectories for all young learners.

While the results are promising, the evidence base is disjointed. A systematic review of the application of AI in special education proposed the diverse applications of the technology but found that methodological limitations were present in all studies (Hopcan et al., 2023). Numerous studies are based on small sample sizes, short follow-ups, diverse outcome measures, and inconsistent parameters of implementation, which are limiting comparability and generalizability across settings (Hopcan et al., 2023). The studies of children with special educational needs in preschool inclusive resource centers also contribute to the discussion of the complexity and changeability of the environment (Kondratyuk et al., 2025). The syntheses that have been done previously tend to focus on feasibility and the benefits in terms of teachers and overlooked child-level outcomes and have not disaggregated the key domains: communication, social interaction, and cognition. Despite new guidelines on the topic of integrating classrooms (Oh-Young & Karlin, 2025), the specific synthesis of AI intervention based on these areas remains absent (Su & Yang, 2022). However, ta lack of evidence exists regarding the best modalities and delivery characteristics (Oh-Young & Karlin, 2025). To address this gap, it is necessary to focus on a strict review that will center on child-level effects in the three areas and evaluate the methodological quality to promote research and practice roadmaps.

In light of the foregoing, this systematic review addresses three questions: 1) what empirical evidence demonstrates the effects of AI-based interventions on communication, social interaction, and cognitive outcomes among young children, primarily in preschool special education contexts; 2) which AI modalities and implementation features—such as delivery model, session dosage, and human-AI orchestration—are associated with differential outcomes in these domains; and 3) what methodological limitations and evidence gaps characterize the current literature.

Given the heterogeneity of the field, this review explicitly distinguishes between studies reporting direct child-level intervention outcomes and those focused on screening, assessment, validation, or system development, to ensure clarity in interpreting evidence of effectiveness.

2. Methods

2.1. Protocols and Reporting Standards

The author developed the protocol a priori and reported the review in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidance to ensure transparent identification, screening, eligibility, and synthesis procedures (Page et al., 2021). The PRISMA 2020 statement represents a revised guideline for reporting systematic reviews that adds more transparency and completeness to reporting. Since the study could not do a quantitative meta-analysis because of the heterogeneity of interventions, outcome measures, and the populations of participants, the author adhered to the Synthesis Without Meta-analysis (SWiM) guidelines in organizing narrative and thematic synthesis and specifying study selection, grouping, and weighting criteria (Campbell et al., 2020). The PICOS approach was used to frame the eligibility criteria, which is a method of supporting transparent and reproducible selection reasoning in health and education reviews (Schardt et al., 2007). The PICOS framework has been confirmed as a useful tool in enhancing the searching of clinical databases to conduct systematic reviews. The review protocol was developed a priori but was not prospectively registered in a public database (e.g., PROSPERO).

2.2. Eligibility Criteria

The selection of the study was based on a preschool-special education PICOS framework. The author selected peer-reviewed studies published in English that reported child-level outcomes among primarily preschool-aged children with developmental disabilities, with inclusion of studies extending into early school-age where relevant, published between January 2010 and January 2026. The significance of early intervention on communication, social, and cognitive development was the reason why autism spectrum disorder, language delay, and global developmental delay were the target conditions. Interventions also demanded an AI element that was directly connected to child development, such as socially assistive robots, adaptive or intelligent tutoring systems, speech-recognition-based applications, augmented or virtual environments, or intelligent agents. Comparators that were eligible included usual instruction, standard therapy, or low-technology assistance. The outcomes were required to respond to at least one of the target domains: communication, social interaction, or cognition and were measured with standardized instruments or structured observational scales. Single-group pre-post designs were acceptable provided that they employed a set of quantitative outcome measures that were capable of tracking a change over time. The studies that only provided qualitative feedback on the feasibility or usability without quantitative results at the level of children were eliminated. These parameters represent the protocol that has been completed regarding screening.

2.3. Information Sources and Search Strategy

An interdisciplinary search was conducted in ERIC, PsycINFO, Scopus, Web of Science, PubMed, and IEEE Xplore databases. Search strings were a combination of intervention terms, population terms, and outcome terms with adaptations to databases in line with PRISMA-S extension requirements of reporting literature searches in systematic reviews (Rethlefsen et al., 2021). The terms used as illustrations were artificial intelligence, machine learning, social robots, intelligent tutoring, early childhood, preschool, kindergarten, social skills, social communication, cognition, and learning outcomes. The database was narrowed down to English-language publications published since January 2010. The final database search was conducted in January 2026. The author reported the plan to achieve good reporting practices of information retrieval of reviews. The search strategy and indexing changes were normalized across the databases to achieve sensitivity with maximum preservation of precision. The preliminary search in the six databases had 1472 records. Complete database-specific search strings are provided in Appendix A. Appendix A contains the complete database search of all the six databases.

2.4. Study Selection

All records were exported into a Microsoft Excel spreadsheet for reference management purposes. The duplicates were determined by sorting by authors, year, and titles and eliminated manually (n = 186), so the final record was 1286 unique records. Study selection was conducted by the author using predefined PICOS criteria. To enhance methodological rigor, a second independent reviewer verified a randomly selected subset (20%) of records at both the title/abstract and full-text screening stages. Discrepancies were resolved through discussion, with consensus achieved in all cases. Inter-rater agreement for the subset was high (percentage agreement = 91%; Cohen's $\kappa = 0.84$), indicating strong reliability of the screening process.

Following title and abstract screening, 412 records were excluded. The remaining 874 articles were retrieved for full-text review. Of these, 843 articles were excluded due to wrong population (n = 287), wrong intervention (n = 198), wrong outcomes (n = 156), wrong study design (n = 112), duplicate reports (n = 54), and lack of data (n = 36). A total of 31 studies met the inclusion criteria and were included in the final synthesis.

2.5. Operational Definitions

To enhance clarity and reproducibility of the synthesis, key coding terms were defined as follows:

AI Modality: The form of artificial intelligence technology used within the

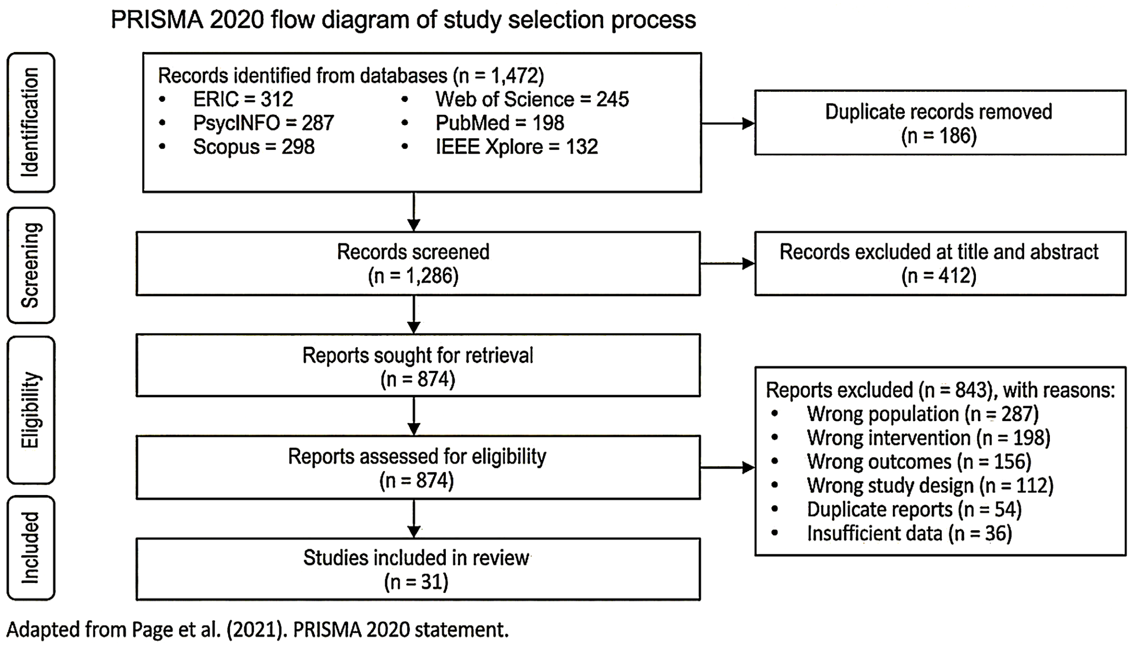

Figure 1. PRISMA 2020 flow diagram for study identification, screening, eligibility, and inclusion.

study (e.g., socially assistive robots, conversational agents, adaptive tutoring systems, augmented/virtual reality platforms).

Implementation Features: Characteristics related to how the intervention was delivered, including session duration, frequency, setting (e.g., home, clinic, classroom), and involvement of adults (e.g., teacher, therapist, caregiver).

Human-AI Orchestration: The manner in which human agents interact with and support AI systems during intervention delivery, including levels of prompting, scaffolding, supervision, and co-participation.

2.6. Data Extraction

A structured data extraction form captured bibliographic details, country, participant characteristics, disability categories, intervention characteristics, AI modality, comparator details, outcome domains, instruments, and key statistics. The data extraction form is provided in Appendix B. Data extraction was conducted by the author using a structured extraction form. To ensure reliability, a second reviewer independently verified a subset (20%) of extracted studies, with discrepancies resolved through discussion. The extraction fields and processes adhered to the Cochrane Handbook of Systematic Reviews of Interventions in the generation of evidence that is reproducible (Cumpston et al., 2019). The summary tables were developed to present the evidence by outcome domain (Table 1) and evidence quality by domain (Table 2). Sample sizes were recorded as reported by study authors; where sample sizes were not reported, this was indicated as “NR” rather than imputed.

2.7. Risk of Bias Assessment

Methodological quality was assessed using design-appropriate tools. For randomized controlled trials (RCTs), the author applied the revised Cochrane Risk of Bias 2 (RoB 2) tool, which assesses bias in 5 domains, including randomization process, deviations in intended interventions, missing outcomes data, measurement of outcomes, and reporting results (Sterne et al., 2019). The RoB 2 tool suggests a different evaluation of risk of bias of randomized trials. In non-randomized trials, such as quasi-experimental designs, single-case experimental designs, and cohort studies, the author relied on Joanna Briggs Institute (JBI) Critical Appraisal Checklists, which touch upon the aspects of confounding, comparability of the participants, measurement validity and reliability, and the appropriateness of the analysis (JBI Manual for Evidence Synthesis, 2020). In the single-case experimental design, standards of the What Works Clearinghouse (WWC) were also used. General scores on low, moderate (some concerns), or high risk of bias were determined. The credibility of effects in synthesis was informed using domain-level judgments. A single, unified master risk of bias table was developed to provide all studies with a single, consistent rating with the same rating that was applied to all tables, appendices, and narrative text. All domain-level quality appraisal results for the 31 studies are presented in Appendix C.

For randomized controlled trials, the RoB 2 domains were defined as follows:

D1 = bias arising from the randomization process;

D2 = bias due to deviations from intended interventions;

D3 = bias due to missing outcome data;

D4 = bias in measurement of outcomes;

D5 = bias in selection of the reported result.

To enable comparison across study designs, a unified overall risk-of-bias rating was applied: studies were classified as low risk when all domains were rated low; moderate risk (“some concerns”) when at least one domain raised concerns; and high risk when one or more domains indicated high risk or when studies lacked sufficient methodological rigor (e.g., feasibility or framework studies without outcome data).

2.8. Synthesis Methods

Given the substantial heterogeneity across AI modalities, implementation parameters, and outcome measures, a meta-analysis was not feasible. Thematic and narrative synthesis was done to combine the findings of communication, social interaction, and cognition. This methodology entailed the use of iterative coding to extract descriptive and explanatory themes associated with effectiveness and context of implementation with special focus on patterns by modality, session structure, and setting. Where the effect sizes and statistical significance were reported, they were extracted to enhance interpretation. The synthesis expressly utilized the methodological quality in balancing the evidence and the other identified consistencies, divergences, and gaps to inform future research (Cumpston et al., 2019). Synthesis of evidence was used to determine patterns within AI technologies, such as social robotics, adaptive tutoring systems, augmented reality platforms, speech recognition tools, and conversational AI in every outcome domain. Table 1 provides the outcome domain summary findings.

Table 1. Summary of evidence by outcome domain.

Domain	Subdomain	n	Studies	Key Findings	Effect Sizes/Statistics	Consistency	Quality
COMMUNICATION	Joint Attention (RJA)	14	Bekele et al., 2013, Cao et al., 2019, 2020, de Belen et al., 2023, Kumazaki et al., 2018, Pérez-Fuster et al., 2022, Scassellati et al., 2018, So et al., 2020, Warren et al., 2015, Zheng et al., 2018	Consistent RJA improvements with robot-mediated interventions. Carryover effects to human interactions demonstrated (50% of ASD children improved JA with humans after robot exposure). Eye-tracking confirmed RJA deficits correlate with cognition and ASD severity. AR-based gaze training achieved 98% PAND.	Kumazaki et al., 2018: $F(1, 26) = 11.45$, $p < 0.01$; Scassellati et al., 2018: $p = 0.001$; de Belen et al., 2023: $F(1, 69) = 11.21$, $p = 0.001$; Pérez-Fuster et al., 2022: PAND 98%, Phi = 0.96	High	Strong
	Joint Attention (IJA)	8	Cao et al., 2019, Ghiglino et al., 2021, So et al., 2020, Annunziata et al., 2024, Baraka et al., 2022, Scassellati et al., 2018	IJA improvements emerged as collateral gains even when not directly targeted. Robot-based intervention superior to human-based for IJA. Initiating social interaction significantly improved with toy robot training.	Ghiglino et al., 2021: $\beta = 4.64$, $t_{23} = 2.39$, $p = 0.026$; So et al., 2020: RBI > HBI (significant); Cao et al., 2019: $F(1, 65) = 11.60$, $p = 0.01$	Moderate	Moderate
	Imitation	7	Bekele et al., 2013, Warren et al., 2015, Zheng et al., 2018, Santos et al., 2021, Annunziata et al., 2024, Baraka et al., 2022, Kostrubiec & Kruck, 2020	ASD children showed greater attention to robot than human (52.76% vs 25.11%). Superior imitation accuracy with robot vs human administrators. RISTA system recognized partial gestures enabling graded feedback. Gesture training feasible with NAO + Kinect.	Bekele et al., 2013: Robot attention 52.76% vs human 25.11%, $p < 0.005$; Warren et al., 2015: ASD imitation robot > human (significant)	Moderate-High	Moderate
	Language/Speech	4	Esfandbod et al., 2023, Fan et al., 2026, Bertacchini et al., 2023, So et al., 2020	Robot-assisted speech therapy (RASA with lip-sync) produced significantly greater language gains than human therapy. Conversational AI	Esfandbod et al., 2023: Overall language $p < .001$, $d = 1.66$; Fan et al., 2026: Initiations MBLR 223% - 598%, Responses MBLR 198%	High	Moderate-Strong

				enhanced peer-mediated intervention dramatically increased initiations and responses. ChatGPT integration enabled real-time adaptive dialogue.	- 9729%, Tau-U = 1.00		
SOCIAL	Emotion Recognition	7	Holeva et al., 2024, Marino et al., 2020, Soleiman et al., 2023, Rudovic et al., 2017, La Fauci De Leo et al., 2025, Zitouni et al., 2026, Rodríguez-Cano et al., 2022	80%-100% emotion recognition accuracy achieved through observational learning with robot-to-robot interactions. Skills maintained and generalized to novel images at 1-month follow-up. Cross-cultural differences in engagement patterns identified. Mood-based recommendations rated beneficial by 100% of participants.	Soleiman et al., 2023: 80% - 100% accuracy maintained at Phase A'; Holeva et al., 2024: NEPSY-II AF improved (both groups); Zitouni et al., 2026: 94% recommendation accuracy	Moderate-High	Moderate
	Social Engagement	10	Clabaugh et al., 2019, Rakhymbayeva et al., 2021, Rudovic et al., 2017, Jürgensen et al., 2026, Kostrubiec & Kruck, 2020, Scassellati et al., 2018, Ghiglino et al. 2021, Baraka et al., 2022	65% average engagement maintained over 41-day in-home deployment with no significant decline. Familiar activities yielded higher engagement than novel ones. Peak engagement associated with positive valence and low arousal. Massive increases in attending (up to 85× baseline).	Clabaugh et al., 2019: 65% engagement, p = 0.99 (no decline); Jürgensen et al., 2026: Tau-U = 0.93 - 1.00, 85× increase (Naeem); Rudovic et al., 2017: r = 0.73 (valence-engagement)	High	Moderate-Strong
	Adaptive/Prosocial Behavior	6	Holeva et al., 2024, Annunziata et al., 2024, Kostrubiec & Kruck, 2020, Scassellati et al., 2018, So et al., 2020, Marino et al., 2020	Robot group showed higher prosocial behavior (SDQ). Training subgroup showed significant ABAS-II Social Adaptive Domain improvement at 6-month follow-up. Caregivers reported increased eye contact and spontaneous communication. Fewer prosocial behaviors with robot vs ball	Holeva et al., 2024: SDQ Prosocial t (42) = 2.457, p = 0.014; Annunziata et al., 2024: ABAS-II SAD p = .022 (at 6 mo); Kostrubiec & Kruck, 2020: Prosocial robot < ball, p = 0.026	Mixed	Moderate

				in low-functioning sample.			
	Perspective-Taking/ToM	3	Holeva et al., 2024, Marino et al., 2020, Rudovic et al., 2017	Both robot-assisted and human-led groups improved on NEPSY-II Theory of Mind. Significant gains in understanding social situations. ToM-based emotion recognition protocol effective across cultures.	Holeva et al., 2024: NEPSY-II ToM improved (both groups, no between-group difference)	Moderate	Moderate
COGNITIVE	Attention/Attending	6	Jürgensen et al., 2026, Clabaugh et al., 2019, Bekele et al., 2013, Zheng et al., 2018, Warren et al., 2015, Rakhymbayeva et al., 2021	Dramatic improvements in attending behavior with SAR (0.4s → 91.4s). Sessions lengthened due to spontaneous social approach behaviors. ASD spent 52.76% time looking at robot vs 25.11% at human. Attention maintained across multi-session protocols.	Jürgensen et al., 2026: Tau-U = 0.759 - 1.00, $p < 0.001$; 85× level increase, 130× frequency increase (Naeem); Bekele et al., 2013: 52.76% vs 25.11%, $p < 0.005$	High	Strong
	Academic/Learning Skills	4	Clabaugh et al., 2019, Lee & Tang, 2024, Zitouni et al., 2026, Romero-García et al., 2021	Significant math gains (numerical operations, reasoning) with in-home SAR. Both robot and teacher instruction improved household task performance vs control; robot slightly better. AI recommender achieved 94% accuracy matching activities to developmental needs. Q-CHAT-NAO validated for screening.	Clabaugh et al., 2019: WIAT-II d = 0.54, $p < 0.01$; Lee & Tang, 2024: Robot and teacher > control; Zitouni et al., 2026: Precision 0.93, Accuracy 0.94	Moderate	Moderate
	Inhibition/Executive Function	2	Holeva et al., 2024, Baraka et al., 2022	Both robot-assisted and human-led groups improved on NEPSY-II Inhibition (fewer errors, less completion time). Therapy mode (personalized sequences) balanced challenge with scaffolding.	Holeva et al., 2024: NEPSY-II IN improved (both groups)	Limited data	Emerging

SCREEN-ING/ASSESS-MENT	Diagnostic Support	3	de Belen et al., 2023, Romero-García et al., 2021, Baraka et al., 2022	Eye-tracking RJA measures correlated with cognitive profiles and ASD severity. 6-item robot-adapted Q-CHAT achieved comparable accuracy to 10-item version. Assess mode established reliable baselines for personalized intervention.	Romero-García et al., 2021: 6-item vs 10-item no significant accuracy loss; de Belen et al., 2023: RJA correlated with MSEL, VABS-II	Moderate	Emerging

Note. RJA = responding to joint attention; IJA = initiating joint attention; ToM = Theory of Mind; SAR = socially assistive robot; AR = augmented reality; PAND = percentage of all non-overlapping data; MBLR = mean baseline reduction; Tau-U = non-parametric effect size for single-case designs. Consistency ratings: High = findings consistent across ≥ 75% of studies; Moderate = consistent across 50% - 74%; Mixed = inconsistent findings; Limited = insufficient studies. Quality ratings based on study designs, risk of bias, sample sizes, and replication across studies. Some studies appear in multiple domains if they measured outcomes across categories.

Table 2. Evidence quality summary by domain.

Evidence Quality	Criteria	Domains
Strong	Multiple RCTs or rigorous SSEDs; consistent findings; large effect sizes; low risk of bias	Joint Attention (RJA), Attention/Attending
Moderate-Strong	At least one RCT plus supportive quasi-experimental evidence; mostly consistent findings	Language/Speech, Social Engagement
Moderate	Multiple studies with some concerns; generally consistent but varied methodology	IJA, Imitation, Emotion Recognition, Adaptive Behavior, Perspective-Taking, Academic Skills
Emerging	Limited studies; preliminary findings; further replication needed	Inhibition/Executive Function, Screening/Assessment

Note. Evidence quality ratings integrate study design rigor, risk of bias assessments, consistency of findings across studies, magnitude of effect sizes, and sample sizes. the most consistent evidence was observed for requires multiple well-designed studies with consistent findings and low risk of bias. RCT = randomized controlled trial; SSED = single-subject experimental design.

3. Results

3.1. Study Selection

The initial database search retrieved 1472 records (ERIC = 312; PsycINFO = 287; Scopus = 298; Web of Science = 245; PubMed = 198; IEEE Xplore = 132). The title and abstract screening was done on 1286 unique records after eliminating 186 duplicates. Out of them, 412 were eliminated due to irrelevance to the review scope, and 874 records were fully assessed in terms of eligibility for full-text review. After careful consideration of the articles in terms of PICOS criteria, 843 articles were excluded due to the following reasons: wrong population (n = 287), wrong intervention (n = 198), wrong outcomes (n = 156), wrong study design (n = 112), duplicate reports (n = 54), and inadequate data (n = 36). Out of the total number of studies, 31 articles were included in the final synthesis because they all fulfilled the inclusion criteria. The flow diagram of PRISMA (**Figure 1**) is the example of the entire screening process. Verification: 1472 − 186 = 1286; 1286 − 412 = 874; 874 − 843 = 31. The further exclusion categories and counts are given in Appendix C.

3.2. Study Characteristics

The 31 included studies were published between 2013 and 2026 and conducted across diverse countries and geographical regions including the United States, Chinese Mainland, Japan, Italy, Spain, Greece, Iran, Kazakhstan, France, Hong Kong SAR, Taiwan Region, South Korea, Belgium, Serbia, Australia, Algeria, Chile, Colombia, Lebanon, and the United Kingdom. The sample sizes were also quite different, including single-case designs consisting of 4 - 6 participants (Jürgensen et al., 2026); Soleiman et al. (2023); Pérez-Fuster et al. (2022) and quasi-experimental cohorts of 193 participants (Lee & Tang, 2024). The target group was mostly preschool-aged children with autism spectrum disorder (ASD), but some studies used children with attention-deficit/hyperactivity disorder (ADHD), developmental delays, language disorders, or mixed groups of typically developing (TD) and neurodiverse children. Detailed study characteristics are presented in Table 3.

Table 3. Characteristics of included studies (N = 31).

Reference	Country and Regions	Design	N	Age	Population
Holeva et al., 2024	Greece	RCT	51	6 - 12 yrs	ASD
Ghiglino et al., 2021	Italy	Crossover	24	~5.8 yrs	ASD
de Belen et al., 2023	Australia	Comparative	77	31 - 73 mo	ASD, TD
Clabaugh et al., 2019	USA	Pre/post	17	3 - 7 yrs	ASD
Cao et al., 2019	Belgium	Mixed	67	~35 - 46 mo	ASD, TD
Cao et al., 2020	Chinese Mainland	Eye-tracking	30	~4 - 5 yrs	ASD, TD
Bertacchini et al., 2023	Italy	Feasibility	NR	NR	ASD
Bekele et al., 2013	USA	Pilot	12	~4 - 5 yrs	ASD, TD
Baraka et al., 2022	Lebanon	Exploratory	11	2 - 7 yrs	ASD
Annunziata et al., 2024	Italy	Pilot	10	50 - 68 mo	ASD
Pérez-Fuster et al., 2022	Spain	SSED	6	3 - 8 yrs	ASD
Lee & Tang, 2024	Taiwan Region	Quasi-exp	193	Preschool	TD
Kumazaki et al., 2018	Japan	RCT	68	5 - 6 yrs	ASD, TD
Kostrubiec & Kruck, 2020	France	Crossover	20	18 - 30 mo	ASD
Fan et al., 2026	Chinese Mainland	Multiple-probe	6	11 - 15 yrs	ASD, ID
Rodríguez-Cano et al., 2022	Chile/Colombia	Framework	18	3 - 15 yrs	ASD

Continued

Santos et al., 2021	Italy	Validation	29	5 - 10 yrs	ASD, TD
Rudovic et al., 2017	Japan/Serbia	Cross-cultural	36	3 - 13 yrs	ASD
Romero-García et al., 2021	Spain	ML validation	1054	Toddlers	ASD
Rakhymbayeva et al., 2021	Kazakhstan	Longitudinal	11	4 - 11 yrs	ASD
So et al., 2020	Hong Kong SAR	Stepped wedge	18	6 - 8 yrs	ASD
Zheng et al., 2018	USA	Feasibility	14	3 - 6 yrs	ASD
Zheng et al., 2016	USA	Comparative	16	~3.8 yrs	ASD, TD
Soleiman et al., 2023	Iran	A-B-A	6	6 - 11 yrs	ASD
Warren et al., 2015	USA	Pilot	16	~3.6 yrs	ASD, TD
So et al., 2020	Hong Kong SAR	Quasi-RCT	38	6 - 9 yrs	ASD
La Fauci De Leo et al., 2025	UK	Framework	NR	NR	ASD
Scassellati et al., 2018	USA	A-B-A	12	6 - 12 yrs	ASD
Zitouni et al., 2026	Algeria	Validation	30	3 - 8 yrs	ASD
Jürgensen et al., 2026	USA	SSED	4	~4.7 yrs	ASD
Esfandbod et al., 2026	Iran	Quasi-RCT	12	6.4 yrs	DSD
Rakhymbayeva et al., 2021	Kazakhstan	Longitudinal	11	4 - 11 yrs	ASD
So et al., 2020	Hong Kong SAR	Stepped wedge	18	6 - 8 yrs	ASD
Zheng et al., 2018	USA	Feasibility	14	3 - 6 yrs	ASD
Zheng et al., 2016	USA	Comparative	16	~3.8 yrs	ASD, TD
Soleiman et al., 2023	Iran	A-B-A	6	6 - 11 yrs	ASD
Warren et al., 2015	USA	Pilot	16	~3.6 yrs	ASD, TD
So et al., 2020	Hong Kong SAR	Quasi-RCT	38	6 - 9 yrs	ASD
La Fauci De Leo et al., 2025	UK	Framework	NR	NR	ASD
Scassellati et al., 2018	USA	A-B-A	12	6 - 12 yrs	ASD
Zitouni et al., 2026	Algeria	Validation	30	3 - 8 yrs	ASD
Jürgensen et al., 2026	USA	SSED	4	~4.7 yrs	ASD
Esfandbod et al., 2023	Iran	Quasi-RCT	12	6.4 yrs	DSD

Intervention characteristics, including technology type, comparator, and duration or dosage, are summarized in Table 4.

Table 4. Intervention characteristics.

Reference	Technology	Comparator	Duration/Dosage
Holeva et al., 2024	NAO robot	Therapist	3 months
Ghiglino et al., 2021	Cozmo robot	Standard	5 weeks
de Belen et al., 2023	Eye-tracking	TD	Single

Continued

Clabaugh et al., 2019	Kiwi SAR	None	41 days
Cao et al., 2019	NAO	Human	Single
Cao et al., 2020	NAO video	Human	Single
Bertacchini et al., 2023	Pepper + AI	None	NR
Bekele et al., 2013	NAO + ARIA	Therapist	30 min
Baraka et al., 2022	NAO	Within	20 min
Annunziata et al., 2024	NAO + Kinect	None	14 weeks
Pérez-Fuster et al., 2022	AR system	Baseline	12 weeks
Lee & Tang, 2024	Kebbi	Teacher	3 weeks
Kumazaki et al., 2018	CommU	Human	3 sessions
Kostrubiec & Kruck, 2020	Prototype	Ball	24 weeks
Fan et al., 2026	Conversational AI	PMI	30 min
Rodríguez-Cano et al., 2022	EMOWY	None	N/A
Santos et al., 2021	NAO + Kinect	None	Variable
Rudovic et al., 2017	NAO	Cross-cultural	25 min
Romero-García et al., 2021	NAO ML	Q-CHAT	N/A
Rakhymbayeva et al., 2021	NAO	None	10 sessions
So et al., 2020	Robot drama	Waitlist	6 sessions
Zheng et al., 2018	NAO system	None	8 weeks
Zheng et al., 2016	NAO system	Human	Multiple
Soleiman et al., 2023	RoboParrot	Baseline	9 sessions
Warren et al., 2015	NAO + Kinect	Human	Single
So et al., 2020	HUMANE	Human	3 weeks
La Fauci De Leo et al., 2025	EmoPal	None	N/A
Scassellati et al., 2018	Autonomous SAR	A-B-A	30 days
Zitouni et al., 2026	YANA	Baseline	Variable
Jürgensen et al., 2026	Kebbi	Human	14 days
Esfandbod et al., 2023	RASA	Human	10 weeks

The outcome measures used across the included studies are presented in Table 5.

Table 5. Outcome measures.

Reference	Outcome Measures
Holeva et al., 2024	CARS-2, NEPSY-II, CBCL
Ghiglino et al., 2021	ESCS
de Belen et al., 2023	ADOS-2, SCQ
Clabaugh et al., 2019	WIAT-II
Cao et al., 2019	RJA, IJA

Continued

Cao et al., 2020	Gaze metrics
Bertacchini et al., 2023	System demo
Bekele et al., 2013	Gaze tracking
Baraka et al., 2022	Task success
Annunziata et al., 2024	ADOS-2
Pérez-Fuster et al., 2022	ESCS
Lee & Tang 2024	Survey
Kumazaki et al., 2018	JA scores
Kostrubiec & Kruck, 2020	Video coding
Fan et al., 2026	Language metrics
Rodríguez-Cano et al., 2022	Questionnaires
Santos et al., 2021	Latency
Rudovic et al., 2017	Engagement
Romero-García et al., 2021	Accuracy
Rakhymbayeva et al., 2021	Engagement
So et al., 2020	ESCS
Zheng et al., 2018	Gaze
Zheng et al., 2016	Gestures
Soleiman et al., 2023	Emotion recognition
Warren et al., 2015	Imitation
So et al., 2020	CARS-2
La Fauci De Leo et al., 2025	Functionality
Scassellati et al., 2018	JA
Zitouni et al., 2026	Accuracy
Jürgensen et al., 2026	Attention

Note. ASD = autism spectrum disorder; TD = typically developing; ADHD = attention-deficit/hyperactivity disorder; ID = intellectual disability; DD = developmental delay; DSD = delayed speech development; GQ = general quotient; DA = developmental age; NVIQ = nonverbal IQ; NR = not reported; M = mean; RCT = randomized controlled trial; SSED = single-subject experimental design; ATD = alternating treatments design; UCD = user-centered design; ML = machine learning; SAR = socially assistive robot; AR = augmented reality; FER = facial expression recognition; ARIA = adaptive robot-mediated intervention architecture; NORRIS = non-contact responsive robot-mediated intervention system; RISTA = robot-mediated imitation skill training architecture; RSE = robotic social environment; CARS-2 = Childhood Autism Rating Scale-Second Edition; NEPSY-II = Developmental Neuropsychological Assessment-Second Edition; AF = Affect Recognition; ToM = Theory of Mind; IN = Inhibition; CBCL = Child Behavior Checklist; TRF = Teacher Report Form; SDQ = Strengths and Difficulties Questionnaire; ESCS = Early Social Communication Scales; SDS = standard difference score; RSDS = restrained standard difference score; RDDS = restrained duration difference score; RT = response time; ADOS-2 = Autism Diagnostic Observation Schedule-Second Edition; SCQ = Social Communication Questionnaire; MSEL = Mullen Scales of Early Learning; VABS-II = Vineland Adaptive Behavior Scales-Second Edition; WIAT-II = Wechsler Individual Achievement Test-Second Edition; RJA = responding to joint attention; IJA = initiating joint attention; LCS = longest common subsequence; LTM = least-to-most; ADI-R = Autism Diagnostic Interview-Revised; ABAS-II = Adaptive Behavior Assessment System-Second Edition; MB-CDI = MacArthur-Bates Communicative Development Inventories; ESES = Educators' Sense of Efficacy Scale; MLU-M = mean length of utterance in morphemes; NDW = number of different words; TNW = total number of words; CLAN = Computerized Language Analysis; FSM = finite state machine; FPFT = first positive feedback time; KBIT-2 = Kaufman Brief Intelligence Test-Second Edition; TOLD = Test of Language Development.

Various AI-mediated interventions were conducted, most of which employed social humanoid robots e.g., NAO (Baraka et al., 2022; Cao et al., 2020; So et al., 2020); CommU (Kumazaki et al., 2018); Pepper (Bertacchini et al., 2023); and Kebbi) and specialized adaptive systems (e.g., ARIA, NORRIS, and RISTA). The average sessions were 25-30 minutes, with the majority of the studies providing 6-21 sessions. The frequency of interventions ranged from single assessment sessions to 30-day daily sessions. Half of the studies involved caregivers, therapists, or teachers. The three studies published in 2026 also used the latest technological innovations, such as conversational AI that can better implement peer-mediated intervention, socially assistive robots to improve attention, and the use of AI-based recommenders of educational activities.

3.3. Thematic Findings

3.3.1. Communication Outcomes

Across studies targeting communication, AI-based tools improved joint attention, imitation, and related foundational behaviors (Pérez-Fuster et al., 2022; So et al., 2020). Numerous robot-mediated methods showed some impressive improvements of joint attention, such as the experiments with autonomous humanoids, which minimized caregiver prompting and maximized triadic interaction (Scassellati et al., 2018; Zheng et al., 2018). According to Zheng et al. (2018) and Scassellati et al. (2018), the results indicated an increase in orienting, shared attention, and early communicative behaviors in preschoolers.

Imitation was a frequent target. Controlled comparisons had indicated greater improvement in imitation and attention under conditions where the practice was mediated by robots as compared to that of human instructors (Warren et al., 2015), and closed-loop or semi-autonomous systems promoted better gesture learning and maintaining attention on tasks (Bekele et al., 2013). An RCT study by Kumazaki et al. (2018) (n = 68 children, 30 ASD, 38 TD) found significant Time x Group interaction in the robot group ($F(1, 26) = 11.45$, $p < 0.01$), with half of the children showing an improvement in joint attention with human partners after interacting with the robot—a carryover effect (F). Bekele et al. (2013) revealed that children with ASD glanced at the robot 52.76 percent of the session time compared to 25.11% when seen by a human therapist ($p < 0.005$), and the ARIA system attained a success rate of 95.83% in the tasks attempted.

Language and speech outcomes were addressed in newer studies. Esfandbod et al. (2023) tested the RASA robot, which supports speech therapy using lip-syncing, with 12 children who have language disorders. There was also a much greater improvement in oral vocabulary ($p = 0.007$), syntactic understanding ($p = 0.007$), and overall language ability ($p < 0.001$) in the robot-assisted group, and the effect size of engagement was d = 1.66. Fan et al. (2026) established that conversational AI-enhanced peer-mediated intervention promoted conversational initiations (MBLR 223% 598) and responses (Tau-U = 1.00) and skills transferred to new interaction partners. The implementation of the augmented reality intervention among six children by Pérez-Fuster et al. (2022) revealed that PAND = 98% in regard to gaze following, and the skills were maintained at a one-month follow-up.

3.3.2. Social Outcomes

Social outcomes were the most extensively studied domain (Holeva et al., 2024). Humanoid robots supported turn-taking, emotion identification, and broader psychosocial engagement (Marino et al., 2020). Holeva et al. (2024) conducted an RCT with 51 children (44 completed) and found that both robot-assisted and human-only therapy yielded significant improvements in CARS-2 scores, NEPSY-II social perception, and inhibition. The only significant between-group difference favored the robot condition on SDQ Prosocial Behavior ($t(42) = 2.457$, $p = 0.014$). Marino et al. (2020) documented significant gains in emotion recognition, perspective-taking, and social situation understanding across 10 sessions with 14 children.

Two studies suggested maintenance and generalization. Soleiman et al. (2023) used a fully robotic social environment where six children observed affective exchanges between two robots. All participants achieved 80%-100% accuracy in emotion recognition, with skills maintained and generalized at one-month follow-up. Annunziata et al. (2024) employed a NAO robot for gesture training in 10 preschoolers with ASD; the subgroup advancing through training (n = 4) showed significant improvement in ABAS-II Social Adaptive Domain at six-month follow-up ($p = 0.022$). So et al. (2023) directly compared robot-based versus human-based interventions for 38 children with high support needs, finding that only the robot-based intervention produced significant and durable improvements in joint attention.

Rudovic et al. (2017) and Ghiglino et al. (2021) conducted a cross-cultural analysis of 36 children from Japan and Serbia and found that engagement was significantly correlated with affective valence-arousal (r = .56-.73). Ghiglino et al. (2021) demonstrated that structured robot activities significantly improved initiation of social interaction ($\beta = 4.64$, $t_{23} = 2.39$, $p = 0.026$). Rakhymbayeva et al. (2021) showed that individualized, adaptive approaches successfully maintained engagement across 7-10 sessions in 11 children with ASD/ADHD, with familiar activities yielding higher engagement than unfamiliar ones.

3.3.3. Cognitive Outcomes

Several studies examined cognitive outcomes including attention, memory, and academic skills. Evidence for cognitive targets such as attention, working memory, problem solving, and early academic skills was positive but more tentative than that for social-communication domains (Clabaugh et al., 2019). Clabaugh et al. (2019) provided 17 children with socially assistive robots in daily in-home sessions over an average of 41 days, observing sustained engagement (65% average) and significant WIAT-II gains (d = 0.54, $p < 0.01$) in numerical operations and math reasoning. Jürgensen et al. (2026) demonstrated dramatic improvements in attending behavior: one participant showed an 85-fold increase in eye contact duration (0.4s → 91.4s), with weighted average Tau-U = 0.930 ($p < 0.001$).

C.-F. Lee & Tang (2024) involved 193 preschoolers comparing AI robot teaching to teacher-led instruction for household tasks. Both instructional groups significantly outperformed the no-instruction control, with the robot-assisted group showing slightly better outcomes. A quasi-experimental study involving an AI teaching robot reported better performance on home assignments than on teacher-led activities (Lee et al., 2022). Zitouni et al. (2026) developed an AI-driven educational activity recommender achieving 94% accuracy in matching activities to developmental needs, with 100% of clinical participants agreeing that mood-based recommendations were beneficial.

Baraka et al. (2022) examined robot action sequencing strategies in 11 children with ASD and found that personalized Therapy mode sequences achieved high task success rates compared to random Explore mode sequences. H.-L. Cao et al. (2020) found that while both ASD and TD children performed better with human partners than robots overall, ADOS severity predicted robot-mediated performance ($R^2 = 0.27$), suggesting that intervention effects may vary by symptom severity. Exploratory implementations using virtual agents or chat-enhanced platforms have demonstrated high engagement (Bertacchini et al., 2023). However, they were limited by small samples, scarce autism-specific data, and the absence of rigorous controls, highlighting the need for larger, theory-driven trials (Bertacchini et al., 2023).

3.3.4. Teacher and Program Support

Several studies examined the use of AI to support screening, assessment, and classroom integration (Dore & Dynia, 2020; Lim & Wardrip, 2024; Nan, 2020; Oh-Young & Karlin, 2025). A robot-administered autism screener produced outcomes aligned with caregiver reports, suggesting the potential for scalable early identification (Romero-García et al., 2021). Short, structured robot activities integrated into preschool classrooms support social communication without disrupting routines (Ghiglino et al., 2021). Eye-tracking paired with robot-supported testing differentiated joint attention profiles associated with cognition and autism severity (De Belen et al., 2023). These tools may alleviate educators' workload and augment program evaluation, although direct child-level gains were mixed (Rodríguez-Cano et al., 2022).

A study-level summary of outcome domains and primary outcomes is provided in **Table 6**.

Table 6. Summary of findings for included studies (N = 31).

Reference	Outcome Domain	Primary Outcomes
Holeva et al., 2024	Social, Cognitive	CARS-2, NEPSY-II, CBCL
Ghiglino et al., 2021	Communication	ESCS
de Belen et al., 2023	Communication	Eye-tracking RJA
Clabaugh et al., 2019	Cognitive	WIAT-II
Cao et al., 2019	Communication	RJA, IJA
Cao et al., 2020	Communication	Gaze metrics
Bertacchini et al., 2023	Cognitive	System function
Bekele et al., 2013	Communication	Gaze tracking
Baraka et al., 2022	Communication, Cognitive	Task success
Annunziata et al., 2024	Social	ABAS-II
Pérez-Fuster et al., 2022	Communication	ESCS
Lee & Tang, 2024	Cognitive	Chores scale
Kumazaki et al., 2018	Communication	JA scores
Kostrubiec & Kruck, 2020	Social	Video coding
Fan et al., 2026	Communication	Language metrics
Rodríguez-Cano et al., 2022	Social	Preferences
Santos et al., 2021	Motor	Movement latency
Rudovic et al., 2017	Social	Engagement
Romero-García et al., 2021	Screening	Accuracy
Rakhymbayeva et al., 2021	Social	Engagement
So et al., 2020	Communication	ESCS
Zheng et al., 2018	Communication	Gaze
Zheng et al., 2016	Communication	Gestures
Soleiman et al., 2023	Social	Emotion recognition
Warren et al., 2015	Communication	Imitation
So et al., 2020	Communication	ESCS
La Fauci De Leo et al., 2025	Social	Functionality
Scassellati et al., 2018	Social	JA
Zitouni et al., 2026	Cognitive	Accuracy
Jürgensen et al., 2026	Cognitive	Attention
Esfandbod et al., 2023	Communication	TOLD

The main findings reported by each included study are summarized in **Table 7**.

Table 7. Key findings.

Reference	Key Findings
Holeva et al., 2024	Both groups improved; robot group higher prosocial behavior
Ghiglino et al., 2021	Robot improved initiation of social interaction
de Belen et al., 2023	TD > ASD in gaze accuracy; slower RT in ASD
Clabaugh et al., 2019	Math improved; engagement stable
Cao et al., 2019	Human > robot for RJA/IJA
Cao et al., 2020	More face attention to robot
Bertacchini et al., 2023	Successful AI integration
Bekele et al., 2013	Higher attention to robot
Baraka et al., 2022	Therapy mode effective
Annunziata et al., 2024	Improvement in adaptive skills
Pérez-Fuster et al., 2022	RJA improved and generalized
Lee & Tang, 2024	Robot slightly better than teacher
Kumazaki et al., 2018	Robot improved JA with carryover
Kostrubiec & Kruck, 2020	Robot reduced human-directed behavior
Fan et al., 2026	Large gains in communication
Rodríguez-Cano et al., 2022	Preference for anthropomorphic robots
Santos et al., 2021	System tracked movement effectively
Rudovic et al., 2017	Engagement varied culturally
Romero-García et al., 2021	6-item tool effective
Rakhymbayeva et al., 2021	Engagement maintained
So et al., 2020	RJA improved significantly
Zheng et al., 2018	Real-time gaze tracking effective
Zheng et al., 2016	Robot improved attention/imitation
Soleiman et al., 2023	Emotion recognition improved
Warren et al., 2015	Higher attention to robot
So et al., 2020	Robot intervention superior
La Fauci De Leo et al., 2025	App functional
Scassellati et al., 2018	JA improved and generalized
Zitouni et al., 2026	High accuracy system
Jürgensen et al., 2026	Large attention gains
Esfandbod et al., 2023	Language improved significantly

As shown in Table 8, the included studies varied in reported effect sizes, statistical indicators, and evidence of maintenance or generalization.

Table 8. Effect sizes and maintenance.

Reference	Statistics (Short)	Maintenance
Holeva et al., 2024	t = 2.457, $p = 0.014$	Post only
Ghiglino et al., 2021	$\beta = 4.64$, $p = 0.026$	Post only
de Belen et al., 2023	F = 11.2, $p = 0.001$	N/A
Clabaugh et al., 2019	d = 0.54, $p < 0.01$	Post
Cao et al., 2019	F = 103, $p < 0.001$	N/A
Cao et al., 2020	$p = 0.038/0.004$	N/A
Bertacchini et al., 2023	N/A	N/A
Bekele et al., 2013	$p < 0.005$	N/A
Baraka et al., 2022	Descriptive	N/A
Annunziata et al., 2024	$p = 0.022$	6-month
Pérez-Fuster et al., 2022	Phi = 0.96	1-month
Lee & Tang, 2024	Significant	1-week
Kumazaki et al., 2018	F = 11.45	Carryover
Kostrubiec & Kruck, 2020	$p < 0.05$	Post
Fan et al., 2026	Tau-U = 1.00	Generalization
Rodríguez-Cano et al., 2022	Scores	N/A
Santos et al., 2021	Latency values	N/A
Rudovic et al., 2017	r = 0.73/0.56	N/A
Romero-García et al., 2021	No diff	N/A
Rakhymbayeva et al., 2021	Significant	Sessions
So et al., 2020	Significant	Maintained
Zheng et al., 2018	Model validated	Sessions
Zheng et al., 2016	Significant	N/A
Soleiman et al., 2023	80% - 100%	Maintained
Warren et al., 2015	Significant	N/A
So et al., 2020	Significant	Delayed
La Fauci De Leo et al., 2025	N/A	N/A
Scassellati et al., 2018	$p = 0.001$	30 days
Zitouni et al., 2026	93% - 94%	N/A
Jürgensen et al., 2026	Tau-U = 0.93	Intervention
Esfandbod et al., 2023	$p < 0.001$	Post

Note. JA = joint attention; RJA = responding to joint attention; IJA = initiating joint attention; MBLR = mean baseline reduction; PAND = percentage of all non-overlapping data; Tau-U = non-parametric effect size for single-case designs; SDS = standard difference score; RSDS = restrained standard difference score; RT = response time; LCS = longest common subsequence; LTM = least-to-most; LTM-RI = least-to-most robot intervention;

FPFT = first positive feedback time; RL = reinforcement learning; NN = neural network; CAI-PMI-ID = conversational AI-enhanced peer-mediated intervention by peers with intellectual disabilities; PMI-ID-only = peer-mediated intervention by peers with intellectual disabilities without AI; RBI = robot-based intervention; HBI = human-based intervention; RAST = robot-assisted speech therapy; NVIQ = nonverbal IQ; CARS = Childhood Autism Rating Scale; SAD = Social Adaptive Domain; GQ = general quotient; ESES = Educators' Sense of Efficacy Scale; MLU-M = mean length of utterance in morphemes; NDW = number of different words; TNW = total number of words; N/A = not applicable.

3.4. Quality Assessment

Methodological quality was heterogeneous. Of the 31 included studies, 11 (35.5%) were rated as low risk of bias, including RCTs by Holeva et al. (2024) and Kumazaki et al. (2018), as well as methodologically rigorous single-case experimental designs meeting WWC standards by Pérez-Fuster et al. (2022), Fan et al. (2026), Soleiman et al. (2023), Scassellati et al. (2018), and Jürgensen et al. (2026). Fifteen studies (48.4%) were rated at moderate risk (some concerns) due to small samples, limited blinding, quasi-randomization, or single-group designs. Five studies (16.1%) were rated at high risk, primarily technical feasibility or design framework studies lacking clinical outcome data (Bertacchini et al., 2023; La Fauci De Leo et al., 2025; Santos et al., 2021). Common methodological limitations included small sample sizes, lack of long-term follow-up, and heterogeneous outcome measures. The strongest evidence supported improvements in joint attention, imitation, and emotion recognition. Cognitive outcomes, while promising, had weaker evidence due to fewer controlled trials. Complete risk of bias assessments are presented in **Table 9**.

Table 9. Risk of bias assessment for included studies; study design, tool and overall risk (N = 31).

Reference	Design	Tool	Overall
Holeva et al., 2024	RCT	RoB 2	Low
Ghiglino et al., 2021	Crossover	RoB 2	Some concerns
de Belen et al., 2023	Comparative	JBI	Low
Clabaugh et al., 2019	Pre/post	JBI	Moderate
Cao et al., 2019	Mixed	JBI	Low
Cao et al., 2020	Eye-tracking	JBI	Low
Bertacchini et al., 2023	Feasibility	N/A	High
Bekele et al., 2013	Pilot	JBI	Moderate
Baraka et al., 2022	Exploratory	JBI	Moderate
Annunziata et al., 2024	Pilot	JBI	Moderate
Pérez-Fuster et al., 2022	SSED	WWC	Low
Lee & Tang, 2024	Quasi-exp	JBI	Some concerns
Kumazaki et al., 2018	RCT	RoB 2	Low

Continued

Kostrubiec & Kruc, 2020	Crossover	RoB 2	Some concerns
Fan et al., 2026	SSED	WWC	Low
Rodríguez-Cano et al., 2022	Framework	N/A	High
Santos et al., 2021	Validation	JBI	High
Rudovic et al., 2017	Cross-cultural	JBI	Some concerns
Romero-García et al., 2021	ML validation	JBI	Moderate
Rakhymbayeva et al., 2021	Longitudinal	JBI	Moderate
So et al., 2020	Stepped wedge	RoB 2	Low
Zheng et al., 2018	Feasibility	JBI	Moderate
Zheng et al., 2016	Comparative	JBI	Some concerns
Soleiman et al.,2023	A-B-A	WWC	Low
Warren et al., 2015	Pilot	JBI	Some concerns
So et al., 2020	Quasi-RCT	RoB 2	Some concerns
La Fauci De Leo et al., 2025	Framework	N/A	High
Scassellati et al., 2018	A-B-A	WWC	Low
Zitouni et al., 2026	Validation	JBI	Some concerns
Jürgensen et al., 2026	SSED	WWC	Low
Esfandbod et al., 2023	Quasi-RCT	RoB 2	Some concerns

Table 10. Risk of bias domains.

Reference	D1	D2	D3	D4	D5
Holeva et al., 2024	Low	Low	Some concerns	Low	Low
Ghiglino et al., 2021	Some concerns	Low	Low	Low	Low
de Belen et al., 2023	Low	Low	Low	Low	Low
Clabaugh et al., 2019	High	Low	Some concerns	Low	Low
Cao et al., 2019	Low	Low	Low	Low	Low
Cao et al., 2020	Low	Low	Low	Low	Low
Bertacchini et al., 2023	N/A	N/A	N/A	N/A	N/A
Bekele et al., 2013	Some concerns	Low	Low	Low	Low
Baraka et al., 2022	Some concerns	Some concerns	Low	Some concerns	Low
Annunziata et al., 2024	High	Low	Low	Low	Low
Pérez-Fuster et al., 2022	Low	Low	Low	Low	Low
Lee & Tang, 2024	Some concerns	Low	Low	Some concerns	Low
Kumazaki et al., 2018	Low	Low	Low	Low	Low
Kostrubiec & Kruck, 2020	Low	Some concerns	Low	Low	Low

Continued

Fan et al., 2026	Low	Low	Low	Low	Low
Santos et al., 2021	High	Low	Low	Low	Low
Rudovic et al., 2017	Some concerns	Low	Low	Low	Low
Romero-García et al., 2021	Low	Low	Low	Low	Low
Rakhymbayeva et al., 2021	Some concerns	Low	Low	Low	Low
So et al., 2020	Low	Low	Low	Low	Low
Zheng et al., 2018	Some concerns	Low	Low	Low	Low
Zheng et al., 2016	Some concerns	Low	Some concerns	Low	Low
Soleiman et al., 2023	Low	Low	Low	Low	Low
Warren et al., 2015	Some concerns	Low	Some concerns	Low	Low
So et al., 2020	Some concerns	Low	Low	Low	Low
La Fauci De Leo et al., 2025	N/A	N/A	N/A	N/A	N/A
Scassellati et al., 2018	Low	Low	Low	Low	Low
Zitouni et al., 2026	Some concerns	Low	Low	Some concerns	Low
Jürgensen et al., 2026	Low	Low	Low	Low	Low
Esfandbod et al., 2023	Some concerns	Low	Low	Low	Low

Notes: RoB 2 = Cochrane Risk of Bias 2 tool; JBI = Joanna Briggs Institute; WWC SSED = What Works Clearinghouse Single-Case Design; N/A = not applicable.

Table 11. Summary of risk bias ratings.

Overall Risk of Bias	n	%	Studies
Low	11	35.5	Holeva et al., 2024, de Belen et al., 2023, Cao et al., 2019, Cao et al., 2020, Pérez-Fuster et al., 2022, Kumazaki et al., 2018, Fan et al., 2026, So et al., 2020, Soleiman et al., 2023, Scassellati et al. 2018, Jürgensen et al., 2026
Some concerns/Moderate	15	48.4	Ghiglino et al. 2021, Clabaugh et al., 2019, Bekele et al., 2013, Baraka et al., 2022, Annunziata et al., 2024, Lee & Tang 2024, Kostrubiec & Kruck, 2020, Rudovic et al., 2017, Romero-García et al., 2021, Rakhymbayeva et al., 2021, Zheng et al., 2018, Zheng et al., 2016, Warren et al., 2015, So et al., 2020, Zitouni et al., 2026, Esfandbod et al., 2023
High	5	16.1	Bertacchini et al., 2023, Rodríguez-Cano et al., 2022, Santos et al., 2021, La Fauci De Leo et al., 2025

Note. Low risk studies included well-designed RCTs, rigorous single-subject experimental designs meeting WWC standards, or comparative studies with objective measures and adequate controls. Moderate risk studies had methodological limitations such as small samples, lack of blinding, pseudo-randomization, high attrition, or absence of control conditions. High risk studies were primarily technical feasibility or design framework papers without clinical outcome data.

3.5. Contextual and Non-Intervention Studies

In addition to intervention studies reporting direct child-level outcomes, several included studies focused on screening, assessment, system validation, or framework development. These included eye-tracking diagnostic tools, machine learning validation studies, and conceptual or feasibility frameworks.

While these studies provide important insights into the broader ecosystem of AI applications in early childhood and special education—particularly in areas such as early detection, progress monitoring, and system design—they do not constitute direct evidence of intervention effectiveness. Accordingly, these studies were analyzed separately and were not included in conclusions regarding intervention efficacy.

4. Discussion

A consistent body of research has demonstrated that AI, with particular emphasis on social robotics, is positively associated with enhancements in joint attention and imitation skills among preschool-aged children with developmental disabilities. Nao-based programs that utilize humanoid robots have demonstrated promising effects in eliciting joint-attention responses (Kumazaki et al., 2018; Scassellati et al., 2018; Zheng et al., 2018). In comparative studies, some studies reported advantages over human-led conditions, while others found no significant differences, particularly for broader social and cognitive outcomes (Warren et al., 2015; Holeva et al., 2024). Gaze following and pointing complementary gains have now been shown in augmented reality and performance-based formats (Pérez-Fuster et al., 2022; So et al., 2020), indicating a close-term possibility of generalization to natural settings. In the social realms, RCTs and pilots showed better emotion recognition, perspective-taking, and social tolerance (Cao et al., 2019; Holeva et al., 2024; Kumazaki et al., 2018), some of which are maintained at follow-up (Annunziata et al., 2024; Soleiman et al., 2023). Furthermore, there is evidence that affective engagement is culturally diverse (Rudovic et al., 2017). The cognitive effects were less convincing but with positive tendencies in attention, engagement, and task performance (Clabaugh et al., 2019; H. Lee et al., 2022), especially when systems were adjusted to child profiles (Bekele et al., 2013; Cao et al., 2020). Teacher and program-facing applications assist in screening, integration of the classroom, and assessment (Ghiglino et al., 2021; Romero-García et al., 2021). These findings are consistent with the thematic synthesis presented in **Table 1**, the study characteristics summarized in **Table 3**, and the methodological quality appraisal presented in **Tables 9-11**.

Importantly, a substantial proportion of the included studies relied on within-group improvements without direct comparison conditions, limiting causal interpretation of effectiveness. Moreover, several studies reported mixed or context-dependent findings, particularly for broader social and cognitive outcomes. Therefore, conclusions regarding effectiveness should be interpreted with caution, and future research employing rigorous controlled designs is needed to establish comparative efficacy.

A central implication of this review is that AI works best when it is a means to a defined pedagogical end rather than a novelty. The highest effects, including those on joint attention and imitation with NAO/NORRIS, are in line with the "tool/tutee" roles in early childhood digital pedagogy (Kostrubiec & Kruck, 2020; Lee et al., 2022), unlike the weaker results of undirected, "software-driven" uses (Lim & Wardrip, 2024). Similar syntheses state as well that technology has a stronger effect when incorporated in the quality improvement models and small, high-value objectives (Ling et al., 2022; Professional Development for Digital Competencies in Early Childhood Education and Care, 2023). Moreover, the effect sizes of proximal measures, which are developed by researchers, are usually bigger compared to broad and standardized outcomes (Oxlad et al., 2021). The key to this promise is personalized affordances (Christakis & Hale, 2025). However, implementation should not be too inhuman (Chen et al., 2022) or too insensitive to differences at the child level (Barry et al., 2025). These interpretations align with the distribution of outcomes and implementation characteristics summarized across **Table 1** and **Table 4**.

These conclusions sit within the wider debates on ethics and capacity building. Key risks, including privacy, prejudice, unfairness, and lack of teacher assistance, are similar to those in early AI (Christakis & Hale, 2025). Most teachers are very unconfident and have little knowledge (Neugnot-Cerioli & Laurenty, 2024), and this argument justifies co-designing professional learning, which creates professional vision (Tammets & Ley, 2023) and maintains human judgment (Matos et al., 2025). Adaptive support is enhanced with targeted training, which is demonstrated empirically (Wullschleger et al., 2023). Conversely, the developmental effect of pedagogy (not the machine) is formed (Su et al., 2024). In line with the general EdTech results, like interactive e-books and AI-assisted bilingual reading (Cao et al., 2019), our data support the conclusion that interactivity regarding a particular learning objective led to results, especially in ASD joint-attention exercises (Zheng et al., 2018). These trends are reflected in the evidence synthesis and intervention characteristics summarized in **Table 1** and **Table 4**.

Social robots, like NAO and CommU, may be utilized in inclusive preschool to act as a predictable, engaging partner that enables children to rehearse basic prerequisites, including joint attention and imitation, while teachers distinguish between instruction or make observations (Warren et al., 2015). Pacing, prompts, and feedback of adaptive tutoring systems and AR platforms can also be customized to the profiles of learners. Notably, socially assistive robots are able to generalize and practice intensively and at high levels, including at home and community levels, facilitating generalization and the ability to support (Cao et al., 2019; Clabaugh et al., 2019) and facilitate structured processes, including CBT elements (Marino et al., 2020), and emotion recognition (Soleiman et al., 2023). Scalable screening and assessment at the systems level (De Belen et al., 2023) and classroom-compatible integration (Ghiglino et al., 2021) can be used to address resource gaps and align with such strategic agendas as the Vision 2030 of Saudi Arabia and UN SDG 4. Additionally, it is important to distinguish between intervention studies and contextual or system-level research. Screening, validation, and feasibility studies—while valuable for informing implementation and scalability—

should not be interpreted as evidence of intervention effectiveness. This distinction is essential for accurately interpreting the current state of the evidence base. Such practice implications are aligned with the characteristics of the study we have synthesized, like the common session pattern of approximately 27.5 minutes provided about 10 times (Table 3), as well as the design and screening procedure summarized in the PRISMA figure (Figure 1). This distinction is further supported by the categorization of study types and outcomes summarized in Table 3, Table 6, and Table 7.

It is imperative that there be a culturally responsive design. Cross-cultural research shows that affect, arousal, and engagement profiles vary by context (Rudovic et al., 2017), highlighting the need for Arabic-language NLP, culturally appropriate simulations, and locally meaningful gestures in intervention design. Such directions are in line with the geographical spread of studies and found gaps (Table 4).

4.1. Strengths and Limitations

The strength of this review is its rigorous, prespecified methodology: PICOS-aligned eligibility, dual screening and extraction, and design-appropriate appraisal (Cochrane RoB 2; JBI). To ensure transparency and replicability, these steps, which are described in the Methods section and illustrated in the PRISMA flow (Figure 1), are applicable. The evidence base of mixed methodology included a variety of modalities of AI (social robotics, adaptive systems, AR/virtual agents) and environments, which facilitated a synthesis of practices (Table 1). The detailed appendices (A-D) give complete disclosure of search methods, extraction methodologies, quality review, and decision to rule out.

Limitations mirror the field's developmental stage: few large RCTs, small samples, brief interventions (often ≈10 sessions), limited to no long-term follow-up, and heterogeneity of measures, which makes synthesis and generalization difficult. The limitation to English-language literature might have left out pertinent studies. These limitations, as indicated by the quality appraisal (Table 4), put the risk of publication bias higher and the overgeneralization of promising initial effects at risk (Bertacchini et al., 2023; So et al., 2020).

4.2. Implications for Practice

AI-based interventions seem to be the most effective when specific, clearly defined skills, e.g., joint attention, imitation, or emotion recognition, are targeted. Sessions of 20 - 30 minutes with a regular schedule (23 times a week) ensure maximum activity. The role of adults will not lose its significance; the use of AI tools must complement human teaching. Response to intervention is predicted by individual differences in baseline skills, which imply the necessity of individual assessment (Rakhymbayeva et al., 2021; So et al., 2023).

4.3. Future Research Directions

The next phase should prioritize: 1) adequately powered, multi-site RCTs with longer duration and follow-up to test retention and cross-context generalization; 2) multilingual, culturally adapted systems, such as Arabic speech/NLP, and culturally aligned social scripts validated with local communities; and 3) co-designed pipelines linking educators, therapists, and developers to embed pedagogical/therapeutic principles from the outset. These directions follow directly from the gaps identified in our synthesis (**Table 4**) and selection overview (**Figure 1**).

5. Conclusion

This 31-study systematic review provides support that AI-based interventions can adequately address the communication, social, and cognitive development of preschool children with developmental disabilities. Joint attention and imitation on the humanoid robot NAO were found to have the best effects when mediated by a robot on a humanoid platform. Social outcomes also became better, such as emotion recognition, perspective-taking, and adaptive social functioning. A minor number of studies have been maintained in follow-up, and there is limited evidence in the long term. There were positive trends in the cognitive outcomes regarding attention, engagement, and task performance, though the evidence base of cognition is weak compared to that of communication and social domains.

Implementation choices mattered: the most effective ones involved sessions of about 20 to 30 minutes, focused on a certain skill, and involved repeated, structured practice as opposed to tool-first application. The role of humans was found to be more effective since the research involving robots with planned scaffolding by adults showed them to be more on-task and more predictable, which suggests that AI is most effective as a supplement to the judgment of an educator or therapist. The screening, classroom integration, and assessment program-support applications were plausible and scalable, but even in child-level gains, direct and goal-focused intervention was necessary.

Small sample sizes, brief periods, and measures were of mixed methodological quality as they were not homogeneous. The risk of bias is moderate, which restricts the accuracy of estimates of effect and necessitates larger and longer and higher-quality studies. The extrapolation of this to non-study environments and other cultures is not yet guaranteed since data in other areas of the world is limited, and culturally responsive design is not well developed. This systematic review finding is consistent with the targeted application of AI with preschoolers with developmental disabilities, the emphasis on the joint attention and imitation objectives, and the integration of tools into a regular routine in school and home; more general assertions of cognitive acceleration, long-term maintenance, and cross-context transfer should be put to a test through more rigorous trials with longer follow-ups and culturally specific Designs.

Ethical Approval

This study was a systematic review of published literature and did not involve human participants, identifiable personal data, or animal subjects. Therefore, formal ethical approval was not required in accordance with institutional and international research guidelines.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

(2020). *JBI Manual for Evidence Synthesis.* JBI. https://doi.org/10.46658/JBIMES-20-01

Amirova, A., Rakhymbayeva, N., Yadollahi, E., Sandygulova, A., & Johal, W. (2021). 10 Years of Human-Nao Interaction Research: A Scoping Review. *Frontiers in Robotics and AI, 8,* Article 744526. https://doi.org/10.3389/frobt.2021.744526

Annunziata, S., Santos, L., Caglio, A., Geminiani, A., Brazzoli, E., Piazza, E. et al. (2024). Interactive Mirroring Games with Social Robot (IOGIOCO): A Pilot Study on the Use of Intransitive Gestures in a Sample of Italian Preschool Children with Autism Spectrum Disorder. *Frontiers in Psychiatry, 15,* Article 1356331. https://doi.org/10.3389/fpsyt.2024.1356331

Baraka, K., Couto, M., Melo, F. S., Paiva, A., & Veloso, M. (2022). "Sequencing Matters": Investigating Suitable Action Sequences in Robot-Assisted Autism Therapy. *Frontiers in Robotics and AI, 9,* Article 784249. https://doi.org/10.3389/frobt.2022.784249

Barry, R., Neumann, M. M., & Neumann, D. L. (2025). Individual Differences and Young Children's Engagement with a Social Robot. *Computers in Human Behavior: Artificial Humans, 4,* Article ID: 100139. https://doi.org/10.1016/j.chbah.2025.100139

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013). A Step Towards Developing Adaptive Robot-Mediated Intervention Architecture (ARIA) for Children with Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 21,* 289-299. https://doi.org/10.1109/tnsre.2012.2230188

Bertacchini, F., Demarco, F., Scuro, C., Pantano, P., & Bilotta, E. (2023). A Social Robot Connected with ChatGPT to Improve Cognitive Functioning in ASD Subjects. *Frontiers in Psychology, 14,* Article 1232177. https://doi.org/10.3389/fpsyg.2023.1232177

Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S. et al. (2020). Synthesis without Meta-Analysis (Swim) in Systematic Reviews: Reporting Guideline. *BMJ, 368,* L6890. https://doi.org/10.1136/bmj.l6890

Cao, H., Simut, R. E., Desmet, N., De Beir, A., Van De Perre, G., Vanderborght, B. et al. (2020). Robot-assisted Joint Attention: A Comparative Study between Children with Autism Spectrum Disorder and Typically Developing Children in Interaction with Nao. *IEEE Access, 8,* 223325-223334. https://doi.org/10.1109/access.2020.3044483

Cao, W., Song, W., Li, X., Zheng, S., Zhang, G., Wu, Y. et al. (2019). Interaction with Social Robots: Improving Gaze toward Face but Not Necessarily Joint Attention in Children with Autism Spectrum Disorder. *Frontiers in Psychology, 10,* Article 1503. https://doi.org/10.3389/fpsyg.2019.01503

Chen, S., Lin, P., & Chien, W. (2022). Children's Digital Art Ability Training System Based on AI-Assisted Learning: A Case Study of Drawing Color Perception. *Frontiers in Psychology, 13,* Article 823078. https://doi.org/10.3389/fpsyg.2022.823078

Christakis, D. A., & Hale, L. (2025). *Handbook of Children and Screens: Digital Media, Development, and Well-Being from Birth Through Adolescence.* Springer. https://doi.org/10.1007/978-3-031-69362-5

Clabaugh, C., Mahajan, K., Jain, S., Pakkar, R., Becerra, D., Shi, Z. et al. (2019). Long-Term Personalization of an In-Home Socially Assistive Robot for Children with Autism Spectrum Disorders. *Frontiers in Robotics and AI, 6,* Article 110. https://doi.org/10.3389/frobt.2019.00110

Cumpston, M., Li, T., Page, M. J., Chandler, J., Welch, V. A., Higgins, J. P. et al. (2019). Updated Guidance for Trusted Systematic Reviews: A New Edition of the Cochrane Handbook for Systematic Reviews of Interventions. *Cochrane Database of Systematic Reviews, 10,* ED000142. https://doi.org/10.1002/14651858.ed000142

de Belen, R. A., Pincham, H., Hodge, A., Silove, N., Sowmya, A., Bednarz, T. et al. (2023). Eye-tracking Correlates of Response to Joint Attention in Preschool Children with Autism Spectrum Disorder. *BMC Psychiatry, 23,* Article No. 211. https://doi.org/10.1186/s12888-023-04585-3

Dore, R. A., & Dynia, J. M. (2020). Technology and Media Use in Preschool Classrooms: Prevalence, Purposes, and Contexts. *Frontiers in Education, 5,* Article 600305. https://doi.org/10.3389/feduc.2020.600305

Eapen, V., Črnčec, R., & Walter, A. (2013). Clinical Outcomes of an Early Intervention Program for Preschool Children with Autism Spectrum Disorder in a Community Group Setting. *BMC Pediatrics, 13,* Article No. 3. https://doi.org/10.1186/1471-2431-13-3

Esfandbod, A., Rokhi, Z., Meghdari, A. F., Taheri, A., Alemi, M., & Karimi, M. (2023). Utilizing an Emotional Robot Capable of Lip-Syncing in Robot-Assisted Speech Therapy Sessions for Children with Language Disorders. *International Journal of Social Robotics, 15,* 165-183. https://doi.org/10.1007/s12369-022-00946-2

Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J., & Dawson, G. (2015). Long-term Outcomes of Early Intervention in 6-Year-Old Children with Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 54,* 580-587. https://doi.org/10.1016/j.jaac.2015.04.005

Fan, J., Zeng, J., & Li, H. (2026). Effects of Conversational Ai-Enhanced Peer-Mediated Intervention by Peers with Intellectual Disabilities on Conversational Skills in Children with ASD. *Research in Developmental Disabilities, 170,* Article ID: 105249. https://doi.org/10.1016/j.ridd.2026.105249

Feng, Y., & Wang, X. (2023). A Comparative Study on the Development of Chinese and English Abilities of Chinese Primary School Students through Two Bilingual Reading Modes: Human-AI Robot Interaction and Paper Books. *Frontiers in Psychology, 14,* Article 1200675. https://doi.org/10.3389/fpsyg.2023.1200675

Ghiglino, D., Chevalier, P., Floris, F., Priolo, T., & Wykowska, A. (2021). Follow the White Robot: Efficacy of Robot-Assistive Training for Children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 86,* Article ID: 101822. https://doi.org/10.1016/j.rasd.2021.101822

Holeva, V., Nikopoulou, V. A., Lytridis, C., Bazinas, C., Kechayas, P., Sidiropoulos, G. et al. (2024). Effectiveness of a Robot-Assisted Psychological Intervention for Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders, 54,* 577-593. https://doi.org/10.1007/s10803-022-05796-5

Hopcan, S., Polat, E., Ozturk, M. E., & Ozturk, L. (2023). Artificial Intelligence in Special Education: A Systematic Review. *Interactive Learning Environments, 31,* 7335-7353. https://doi.org/10.1080/10494820.2022.2067186

Jürgensen, M., Kubina, R. M., Leitao, N. D. L., & Winn, K. (2026). Using a Socially Assistive Robot to Improve Attention in Children with Autism during Academic Instruction. *International Journal of Social Robotics, 18,* Article No. 40.

https://doi.org/10.1007/s12369-026-01380-4

Kondratyuk, Z., Omelchenko, I., Konopliasta, S., Matiushchenko, I., Synytsia, A., & Mykhailenko, V. (2025). Adaptive Capacity of Preschoolers with Special Educational Needs in Inclusive Resource Centres. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience, 16,* 41-55. https://doi.org/10.70594/brain/16.s1/4

Kostrubiec, V., & Kruck, J. (2020). Collaborative Research Project: Developing and Testing a Robot-Assisted Intervention for Children with Autism. *Frontiers in Robotics and AI, 7,* Article 37. https://doi.org/10.3389/frobt.2020.00037

Kumazaki, H., Yoshikawa, Y., Yoshimura, Y., Ikeda, T., Hasegawa, C., Saito, D. N. et al. (2018). The Impact of Robotic Intervention on Joint Attention in Children with Autism Spectrum Disorders. *Molecular Autism, 9,* Article No. 46. https://doi.org/10.1186/s13229-018-0230-8

La Fauci De Leo, A., Bagheri Zadeh, P., Voderhobli, K., & Sheikh Akbari, A. (2025). A New AI Framework to Support Social-Emotional Skills and Emotion Awareness in Children with Autism Spectrum Disorder. *Computers, 14,* Article 292. https://doi.org/10.3390/computers14070292

Lee, C., & Tang, S. (2024). Can AI Robot Teaching Improve Children's Performance in Housework? A Quasi-Experimental Study. *Children, 11,* Article 1330. https://doi.org/10.3390/children11111330

Lee, H., Chung, M. A., Kim, H., & Nam, E. W. (2022). The Effect of Cognitive Function Health Care Using Artificial Intelligence Robots for Older Adults: Systematic Review and Meta-Analysis. *JMIR Aging, 5,* e38896. https://doi.org/10.2196/38896

Lim, J., & Wardrip, P. (2024). Technology Integration as a Spectrum: Integrating Technology in Early Childhood Classrooms. *Teachers and Teaching, 32,* 289-307. https://doi.org/10.1080/13540602.2024.2420137

Ling, L., Yelland, N., Hatzigianni, M., & Dickson-Deane, C. (2022). The Use of Internet of Things Devices in Early Childhood Education: A Systematic Review. *Education and Information Technologies, 27,* 6333-6352. https://doi.org/10.1007/s10639-021-10872-x

Marino, F., Chilà, P., Sfrazzetto, S. T., Carrozza, C., Crimi, I., Failla, C. et al. (2020). Outcomes of a Robot-Assisted Social-Emotional Understanding Intervention for Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 50,* 1973-1987. https://doi.org/10.1007/s10803-019-03953-x

Matos, T., Santos, W., Zdravevski, E., Coelho, P. J., Pires, I. M., & Madeira, F. (2025). A Systematic Review of Artificial Intelligence Applications in Education: Emerging Trends and Challenges. *Decision Analytics Journal, 15,* Article ID: 100571. https://doi.org/10.1016/j.dajour.2025.100571

Nan, J. (2020). Research of Application of Artificial Intelligence in Preschool Education. *Journal of Physics: Conference Series, 1607,* Article ID: 012119. https://doi.org/10.1088/1742-6596/1607/1/012119

Neugnot-Cerioli, M., & Laurenty, O. M. (2024). *The Future of Child Development in the AI Era. Cross-Disciplinary Perspectives between AI and Child Development Experts.* arXiv: 2405.19275.

Oh-Young, C., & Karlin, M. (2025). Artificial Intelligence . . . In the Early Childhood Special Education Classroom!!? *TEACHING Exceptional Children, 57,* 348-356. https://doi.org/10.1177/00400599241231237

Oxlad, M., Modde, A., & Turnbull, D. (2021). Learning About, from and with Other Professions: A Systematic Review of Interprofessional Education within Psychology Teaching Programs. *Educational Research Review, 34,* Article ID: 100410. https://doi.org/10.1016/j.edurev.2021.100410

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D. et al. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ, 372,* n71. https://doi.org/10.1136/bmj.n71

Pérez-Fuster, P., Herrera, G., Kossyvaki, L., & Ferrer, A. (2022). Enhancing Joint Attention Skills in Children on the Autism Spectrum through an Augmented Reality Technology-Mediated Intervention. *Children, 9,* Article 258. https://doi.org/10.3390/children9020258

Rakhymbayeva, N., Amirova, A., & Sandygulova, A. (2021). A Long-Term Engagement with a Social Robot for Autism Therapy. *Frontiers in Robotics and AI, 8,* Article 669972. https://doi.org/10.3389/frobt.2021.669972

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J. et al. (2021). PRISMA-S: An Extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Systematic Reviews, 10,* Article No. 39. https://doi.org/10.1186/s13643-020-01542-z

Rodríguez-Cano, S., Cuesta-Gómez, J. L., Delgado-Benito, V., & Fuente-Anuncibay, R. d. l. (2022). Educational Technology as a Support Tool for Students with Specific Learning Difficulties—Future Education Professionals' Perspective. *Sustainability, 14,* Article 6177. https://doi.org/10.3390/su14106177

Romero-García, R., Martínez-Tomás, R., Pozo, P., de la Paz, F., & Sarriá, E. (2021). Q-CHAT-NAO: A Robotic Approach to Autism Screening in Toddlers. *Journal of Biomedical Informatics, 118,* Article ID: 103797. https://doi.org/10.1016/j.jbi.2021.103797

Rudovic, O., Lee, J., Mascarell-Maricic, L., Schuller, B. W., & Picard, R. W. (2017). Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study. *Frontiers in Robotics and AI, 4,* Article 36. https://doi.org/10.3389/frobt.2017.00036

Santos, L., Geminiani, A., Schydlo, P., Olivieri, I., Santos-Victor, J., & Pedrocchi, A. (2021). Design of a Robotic Coach for Motor, Social and Cognitive Skills Training toward Applications with ASD Children. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29,* 1223-1232. https://doi.org/10.1109/tnsre.2021.3091320

Scassellati, B., Boccanfuso, L., Huang, C., Mademtzi, M., Qin, M., Salomons, N. et al. (2018). Improving Social Skills in Children with ASD Using a Long-Term, In-Home Social Robot. *Science Robotics, 3,* eaat7544. https://doi.org/10.1126/scirobotics.aat7544

Schardt, C., Adams, M. B., Owens, T., Keitz, S., & Fontelo, P. (2007). Utilization of the PICO Framework to Improve Searching Pubmed for Clinical Questions. *BMC Medical Informatics and Decision Making, 7,* Article No. 16. https://doi.org/10.1186/1472-6947-7-16

So, W., Cheng, C., Lam, W., Huang, Y., Ng, K., Tung, H. et al. (2020). A Robot-Based Play-Drama Intervention May Improve the Joint Attention and Functional Play Behaviors of Chinese-Speaking Preschoolers with Autism Spectrum Disorder: A Pilot Study. *Journal of Autism and Developmental Disorders, 50,* 467-481. https://doi.org/10.1007/s10803-019-04270-z

So, W., Law, W., Cheng, C., Lee, C., Ng, K., Kwok, F. et al. (2023). Comparing the Effectiveness of Robot-Based to Human-Based Intervention in Improving Joint Attention in Autistic Children. *Frontiers in Psychiatry, 14,* Article 1114907. https://doi.org/10.3389/fpsyt.2023.1114907

Soleiman, P., Moradi, H., Mehralizadeh, B., Ameri, H., Arriaga, R. I., Pouretemad, H. R. et al. (2023). Fully Robotic Social Environment for Teaching and Practicing Affective Interaction: Case of Teaching Emotion Recognition Skills to Children with Autism Spectrum Disorder, a Pilot Study. *Frontiers in Robotics and AI, 10,* Article 1088582. https://doi.org/10.3389/frobt.2023.1088582

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I. et al. (2019). RoB 2: A Revised Tool for Assessing Risk of Bias in Randomised Trials. *BMJ, 366,* L4898. https://doi.org/10.1136/bmj.l4898

Su, J., & Yang, W. (2022). Artificial Intelligence in Early Childhood Education: A Scoping Review. *Computers and Education: Artificial Intelligence, 3,* Article ID: 100049. https://doi.org/10.1016/j.caeai.2022.100049

Su, J., Yang, W., Yim, I. H. Y., Li, H., & Hu, X. (2024). Early Artificial Intelligence Education: Effects of Cooperative Play and Direct Instruction on Kindergarteners' Computational Thinking, Sequencing, Self-regulation and Theory of Mind Skills. *Journal of Computer Assisted Learning, 40,* 2917-2925. https://doi.org/10.1111/jcal.13040

Tamaral, C., Hernandez, L., Baltasar, C., & Martin, J. S. (2025). Design Techniques for the Optimal Creation of a Robot for Interaction with Children with Autism Spectrum Disorder. *Machines, 13,* Article 67. https://doi.org/10.3390/machines13010067

Tammets, K., & Ley, T. (2023). Integrating AI Tools in Teacher Professional Learning: A Conceptual Model and Illustrative Case. *Frontiers in Artificial Intelligence, 6,* Article 1255089. https://doi.org/10.3389/frai.2023.1255089

Telisheva, Z., Amirova, A., Rakhymbayeva, N., Zhanatkyzy, A., & Sandygulova, A. (2022). The Quantitative Case-By-Case Analyses of the Socio-Emotional Outcomes of Children with ASD in Robot-Assisted Autism Therapy. *Multimodal Technologies and Interaction, 6,* Article 46. https://doi.org/10.3390/mti6060046

Walan, S. (2025). Primary School Students' Perceptions of Artificial Intelligence—For Good or Bad. *International Journal of Technology and Design Education, 35,* 25-40. https://doi.org/10.1007/s10798-024-09898-2

Warren, Z., Zheng, Z., Das, S., Young, E. M., Swanson, A., Weitlauf, A. et al. (2015). Brief Report: Development of a Robotic Intervention Platform for Young Children with ASD. *Journal of Autism and Developmental Disorders, 45,* 3870-3876. https://doi.org/10.1007/s10803-014-2334-0

Wullschleger, A., Lindmeier, A., Heinze, A., Meier-Wyder, A., Leuchter, M., Vogt, F. et al. (2023). Improving the Quality of Adaptive Learning Support Provided by Kindergarten Teachers in Play-Based Mathematical Learning Situations. *European Early Childhood Education Research Journal, 31,* 225-242. https://doi.org/10.1080/1350293x.2022.2081348

Zheng, Z., Zhao, H., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., & Sarkar, N. (2018). Design, Development, and Evaluation of a Noninvasive Autonomous Robot-Mediated Joint Attention Intervention System for Young Children with ASD. *IEEE Transactions on Human-Machine Systems, 48,* 125-135. https://doi.org/10.1109/thms.2017.2776865

Zheng, Z., Young, E. M., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., & Sarkar, N. (2016). Robot-mediated Imitation Skill Training for Children with Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24*, 682-691. https://doi.org/10.1109/tnsre.2015.2475724

Zitouni, H., Bouteldja, F., Tiri, Z., Meshoul, S., & Bensalem, I. (2026). AI-Driven Educational Activity Recommender for Children with Autism. *Applied Sciences, 16,* Article 2386. https://doi.org/10.3390/app16052386