



OPEN ACCESS

EDITED BY

Sara Uccella,
University of Genoa, Italy

REVIEWED BY

Hafiz Arslan Ramzan,
National University of Sciences and
Technology (NUST), Pakistan
Jayanti Dansana,
KIIT University, India

*CORRESPONDENCE

Fatih Gül
✉ fatih.gul@erdogan.edu.tr

RECEIVED 18 June 2026

REVISED 17 July 2026

ACCEPTED 04 August 2026

PUBLISHED 20 August 2026

CITATION

Arslan A and Gül F (2026)
Artificial-intelligence–based
developmental screening for pre-school
children (ages 3–6): a scoping review of
deployed tools, clinical validity, equity,
and implementation in guidance-center
settings. *Front. Psychol.* 17:1913332.
doi: 10.3389/fpsyg.2026.1913332

COPYRIGHT

© 2026 Arslan and Gül. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Artificial-intelligence–based developmental screening for pre-school children (ages 3–6): a scoping review of deployed tools, clinical validity, equity, and implementation in guidance-center settings

Adem Arslan¹ and Fatih Gül^{2*}

¹Department of Child Care and Youth Services, Vocational School of Health Services, Gümüşhane University, Gümüşhane, Türkiye, ²Artificial Intelligence and Internet of Things Research Laboratory, Department of Electrical-Electronics Engineering, Recep Tayyip Erdoğan University, Rize, Türkiye

Background: Early identification of neurodevelopmental conditions during the pre-school years (ages 3–6) is critical for timely intervention, yet traditional questionnaire-based screeners suffer from low positive predictive value, late detection, and inequitable access. Artificial intelligence (AI) and machine-learning (ML) methods promise more objective, scalable identification, and two devices (Cognoa Canvas Dx, 2021; EarliPoint Evaluation, 2023) are already US Food and Drug Administration (FDA)-authorized. Recent reviews summarize the technical literature, but few address how deployed tools perform on clinical validity, equity, and real-world implementation, particularly outside high-income, English-speaking settings.

Methods: Following Joanna Briggs Institute methodology and the PRISMA extension for scoping reviews (PRISMA-ScR), we searched seven databases (1 January 2020–15 June 2026) and charted primary validation studies of AI/ML developmental-screening tools for children aged 72 months or younger. We mapped tools by modality, diagnostic performance, regulatory status, geographic origin, and, as a structured analytic axis, demographic subgroup (equity) reporting and deployment readiness.

Results: Sixteen primary AI-tool studies met inclusion, spanning eight modality classes (eye-tracking, multimodal questionnaire-plus-video, computer-vision home/clinical video, tablet/smartphone digital phenotyping, voice/acoustic ML, electronic-health-record risk prediction, questionnaire-based ML, and wearable sensors) and eight countries (USA, South Korea, Australia, Switzerland, France, Israel, China, Vietnam). Reported sensitivities ranged from approximately 75 to 99.2% and specificities from 78.9 to 95.5%. Only one tool demonstrated stable performance across sex, race, and ethnicity; subgroup-stratified validation was otherwise near-absent. No CE-marked autism-specific screening device was identified.

Conclusions: The evidence base has broadened in modality and geography but remains autism-concentrated and equity-blind, with a wide gap between analytic accuracy and primary-care deployment. We propose equity validation,

deployment-readiness reporting, and an implementation pathway for low-resource guidance-center settings as priorities.

KEYWORDS

algorithmic equity, artificial intelligence, autism spectrum disorder, developmental screening, digital phenotyping, implementation science, machine learning, pre-school children

1 Introduction

Neurodevelopmental conditions, including autism spectrum disorder (ASD), language delay, attention-deficit/hyperactivity disorder (ADHD) precursors, and global developmental delay, affect a substantial fraction of pre-school-aged children. The United States Centers for Disease Control and Prevention's Autism and Developmental Disabilities Monitoring (ADDM) Network reported an ASD prevalence of 1 in 36 children aged 8 years for surveillance year 2020 (Maenner et al., 2023). Earlier identification translates into earlier intervention and improved cognitive, language, and adaptive outcomes (Zwaigenbaum et al., 2015), and the American Academy of Pediatrics recommends universal ASD screening at the 18- and 24-month visits with ongoing surveillance through the pre-school years (Hyman et al., 2020).

For more than two decades, the dominant primary-care screeners have been parent-completed questionnaires, of which the Modified Checklist for Autism in Toddlers–Revised with Follow-up (M-CHAT-R/F) is most widely deployed. In its pivotal validation in over 16,000 toddlers it achieved sensitivity 0.85 and specificity 0.99, but positive predictive value (PPV) was only 0.48 (Robins et al., 2014): roughly half of positive screens are false positives in low-prevalence settings. The screening-to-diagnosis pipeline also breaks down at referral, with average diagnosis ages exceeding 4 years in many regions and well-documented disparities by race, ethnicity, sex, and socioeconomic status (Pham et al., 2022; Aylward et al., 2021; Cruz et al., 2024).

Against this backdrop, AI and ML approaches, encompassing computer vision, deep learning, eye-tracking analytics, voice processing, tablet/smartphone digital phenotyping, wearable sensing, and electronic-health-record (EHR) risk prediction, have emerged as complementary screening modalities. Two devices hold FDA authorization: Cognoa Canvas Dx (De Novo, June 2021) (Megerian et al., 2022) and EarliPoint (510(k), June 2023) (Jones et al., 2023). Numerous research-stage tools report strong diagnostic-accuracy metrics across toddlers and pre-schoolers (Perochon et al., 2023; Ko et al., 2023; Barbaro and Yaari, 2020).

1.1 Why another review, and what this one adds

The AI-autism literature is now actively reviewed. Recent syntheses aggregate dozens to hundreds of research studies and report pooled accuracy ranges: modality-bounded scoping reviews of face/voice/text (Mohammadi et al., 2025) and of voice/behavioral

data (Rakotomanana and Rouhafzay, 2025); a systematic review of biomarkers and explainable AI (Agrawal and Agrawal, 2025); a generative-AI scoping review (Sohn et al., 2025); broad narrative sweeps (Shen and Yu, 2025; Zhang, 2025); a fetal-to-adolescent pediatric-AI systematic review (Wang et al., 2025); a child-development monitoring review with an acceptance lens (Reinhart et al., 2024); and a focused review of smartphone apps for ASD screening (Reddy et al., 2024). A narrative review in this field already covers eye-tracking, voice, and video tools and discusses scalability, regulation, and ethics (Patel and others, 2025), and an unreviewed pre-print specifically targets the pre-school band (Hasan, 2025).

We therefore do not claim primacy on the pre-school focus or on cataloging AI methods. Instead, this review is distinguished by three axes that prior syntheses leave largely untouched (Table 1). First, we profile deployed and near-deployed tools as case studies and assess their deployment readiness rather than aggregating method papers on accuracy alone. Second, we treat equity, demographic subgroup-stratified validation, as a first-class, tabulated analytic axis. Third, we situate adoption in a low-resource guidance-center health system, using Türkiye's Guidance and Research Center (Rehberlik Araştırma Merkezi, RAM) network as a concrete implementation exemplar, a context absent from every prior review we identified.

The primary research question is: what is the current landscape of deployed and research-stage AI-based developmental-screening tools for children aged 72 months or younger, with respect to modality, diagnostic performance, regulatory status, demographic equity, and implementation readiness?

2 Materials and methods

2.1 Protocol, eligibility, and sources

This scoping review follows Joanna Briggs Institute methodology and the PRISMA-ScR checklist (Tricco et al., 2018); the protocol was registered on the Open Science Framework prior to charting. Inclusion criteria were (i) publication between 1 January 2020 and 15 June 2026; (ii) participants aged 72 months or younger at study entry; (iii) an explicit AI/ML/deep-learning screening or risk-prediction component; (iv) English-language peer-reviewed primary studies; and (v) reporting of at least one diagnostic-accuracy metric (sensitivity, specificity, AUC, PPV, NPV) or a prospective real-world deployment outcome. Reviews, regulatory documents, and foundational epidemiological/guideline references were retained separately for context and comparison but were not charted as primary tool studies. Exclusion criteria

TABLE 1 Positioning against recent reviews of AI for autism / developmental screening (✓, addressed substantively; P, partial/in passing; —, not addressed).

Review	Type	Focus	#st.	Adopt.	Equity	Regul.	LMIC
Patel 2025 (Patel and others, 2025)	Narrative	Eye-track/voice/video, ASD/NDD	—	P	P	P	—
Hasan 2025 (Hasan, 2025) [†]	Systematic	Preschool ASD, AI	—	—	—	—	—
Zhang 2025 (Zhang, 2025)	Narrative	Screening+dx+intervention	—	P	—	—	—
Mohammadi 2025 (Mohammadi et al., 2025)	Scoping	Face/voice/text	23	—	P	—	—
Rakotomanana 2025 (Rakotomanana and Rouhafzay, 2025)	Scoping	Voice/behavioral	158	—	P	—	—
Sohn 2025 (Sohn et al., 2025)	Scoping	Generative AI / LLMs	10	—	—	—	—
Agrawal 2025 (Agrawal and Agrawal, 2025)	Systematic	Biomarkers / XAI	106	—	—	—	—
Shen 2025 (Shen and Yu, 2025)	Narrative	Broad technology sweep	—	P	P	—	—
Wang 2025 (Wang et al., 2025)	Systematic	Pediatric AI, fetal–18	133	P	—	—	—
Reinhart 2024 (Reinhart et al., 2024)	Systematic	Child-dev monitoring, acceptance	—	✓	—	—	—
This review	Scoping	deployed tools, eight modalities	16	✓	✓	✓	✓

No prior synthesis combines deployed-tool case studies, a structured equity matrix, a deployment-readiness lens, and a low-resource guidance-center implementation pathway. [†]Non-peer-reviewed pre-print. #st. = number of studies covered (— where not a count-based synthesis). Values in bold denote the present review.

covered conference abstracts without a full paper, editorials/letters, method/benchmark papers on pre-labeled datasets without a target-population validation, adult or school-aged-only cohorts, and non-peer-reviewed pre-prints. Seven databases were searched (PubMed/MEDLINE, Web of Science Core Collection, Scopus, IEEE Xplore, ERIC, Cochrane CENTRAL, [ClinicalTrials.gov](#)), supplemented by citation searching of landmark papers and a hand search of the FDA 510(k)/De Novo database (product code QPF). Full search strings appear in the [Supplementary material](#).

The restriction to English-language reports was pragmatic rather than principled: the review team could not reliably screen, chart, and appraise full texts in languages outside its competence, and no translation resources were available for this unfunded synthesis; in addition, validation studies of AI screening tools are pre-dominantly indexed in English in the seven databases searched. We did not apply this criterion because we regard non-English evidence as less valuable, and we consider the resulting language bias, together with its likely direction, explicitly in the Limitations.

2.2 Selection, charting, and synthesis

Records were de-duplicated and screened independently by both authors; full texts were retrieved for all “include”/“maybe” decisions. A charting form captured citation metadata, tool name, modality, target condition, country, sample size and age range, reference standard, diagnostic-accuracy metrics, regulatory status, demographic subgroup reporting, external-validation status, and identified deployment barriers. A descriptive synthesis was produced, organized by modality, performance, equity, and deployment readiness; no formal meta-analysis or critical appraisal was conducted, consistent with PRISMA-ScR expectations.

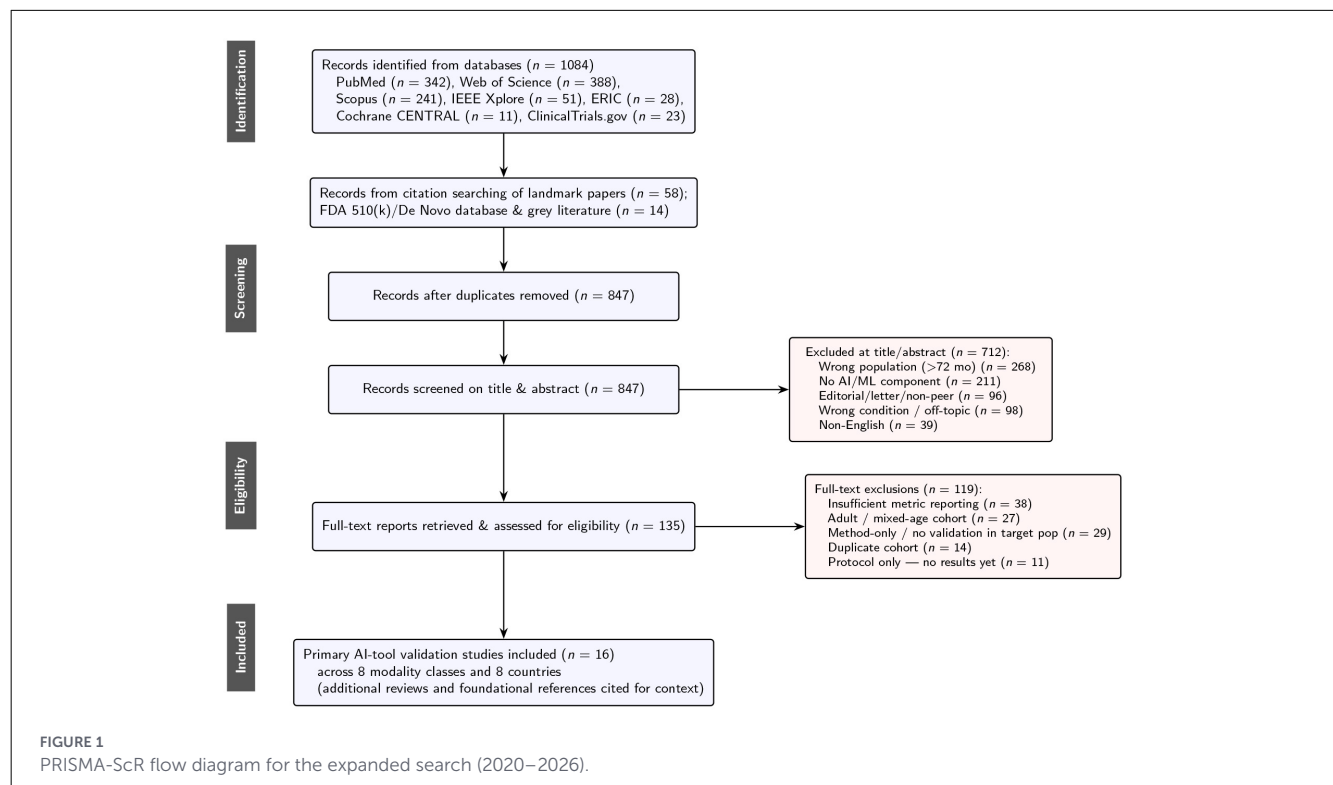
3 Results

3.1 Study selection and landscape

The PRISMA-ScR flow ([Figure 1](#)) summarizes selection. Of 1,084 database records plus 72 from citation/gray sources, 847 remained after de-duplication; 712 were excluded on title/abstract and 135 assessed in full text, yielding 16 primary AI-tool validation studies across eight modality classes and eight countries ([Table 2](#)). All charted primary studies targeted ASD or ASD risk; no primary study addressed language-only, ADHD-only, or global-developmental-delay-only screening at comparable maturity.

3.2 Modalities, performance, and geographic breadth

The 16 tools span eight modalities ([Figure 2](#) and [Table 2](#)). Eye-tracking is represented by the FDA-cleared EarliPoint ([Jones et al., 2023](#)), the 2-min Gazefinder protocol ([Chetcuti et al., 2024](#)), and novel gaze-metric systems ([Wang et al., 2024](#)). Multimodal designs include the FDA-authorized Canvas Dx (caregiver questionnaire plus home video plus clinician inputs) ([Megerian et al., 2022](#)) and a video–audio neural-network ensemble ([Natraj et al., 2024](#)). Computer-vision behavioral analysis appears as joint-attention deep learning ([Ko et al., 2023](#)), automated home-video identification ([Kim et al., 2025](#)), and stereotyped-movement detection ([Barami et al., 2024](#)). Tablet/smartphone digital phenotyping is represented by SenseToKnow ([Perochon et al., 2023](#); [Aikat et al., 2025](#)) and the parent-administered ASDetect app ([Barbaro and Yaari, 2020](#)). Newer modality classes, absent from our earlier mapping, include voice/acoustic classification ([Briend et al., 2023](#)), large-scale



EHR risk prediction (Ben-Sasson et al., 2024), questionnaire-based ML on abbreviated instruments (Pan et al., 2025; Lu et al., 2024; Van Vo et al., 2026), and wearable movement sensing (Wilson et al., 2024).

Reported sensitivities ranged from approximately 75% [well-baby EHR model (Ben-Sasson et al., 2024)] to 99.2% (joint-attention DL Ko et al., 2023), and specificities from 78.9% [Canvas Dx (Megerian et al., 2022)] to 95.5% (joint-attention DL). The highest AUC was 0.996 (Ko et al., 2023), although that system was validated in a single center with 95 children, limiting generalizability. PPV was consistently the weak metric under realistic prevalence, mirroring M-CHAT-R/F (Robins et al., 2014); Figure 3 positions the tools that report both metrics against this benchmark.

These ranges must not be read as a head-to-head comparison of tool performance. The included studies differ fundamentally in design (clinically-referred case-control samples vs. prospective community screening), reference standard (full multidisciplinary diagnostic assessment vs. registry or chart-derived diagnosis), age band, sampling frame, and underlying prevalence—all of which materially shape sensitivity, specificity, and especially PPV. A case-control design that contrasts already-diagnosed autistic children with typically-developing controls will systematically report higher apparent accuracy than the same tool applied to an unselected community population, an effect known as spectrum bias. Consequently, the differences visible in Table 2 and Figure 3 reflect study design at least as much as intrinsic tool quality, and the figures should be read as landscape descriptors rather than as a ranking. This pattern is itself informative: the highest accuracies came from small, single-site, case-control designs, whereas

the largest (Ben-Sasson et al., 2024) ($n = 780,610$) and the most prospectively deployed (Perochon et al., 2023) studies reported more modest, more realistic figures. Geographically, the evidence base now extends well beyond the United States, adding South Korea, Australia, Switzerland, France, Israel, China, and Vietnam, though sub-Saharan Africa, South Asia, the Middle East, and Türkiye remain unrepresented by any primary tool study (Figure 4).

3.3 Equity and subgroup validation

Demographic subgroup reporting was the single weakest dimension across the corpus (Table 3). Only SenseToKnow explicitly reported stable accuracy across sex, race, and ethnicity (Perochon et al., 2023). The well-baby EHR model reported sex-stratified performance (boys-only AUC 0.82) (Ben-Sasson et al., 2024), and EarliPoint's replication cohort was more demographically diverse than its discovery cohort, though per-group AUCs were not reported (Jones et al., 2023). The remaining tools provided either unstratified demographics or none. This matters because diagnostic disparities by race, ethnicity, sex, and socioeconomic status are well-documented (Pham et al., 2022; Aylward et al., 2021), female autism is systematically under-detected (Cruz et al., 2024), and the clinical-AI fairness literature converges on inadequate subgroup-stratified validation as the central unmet need (Liu et al., 2025; Huang et al., 2022). An AI screener trained and validated without stratification risks encoding and amplifying these disparities.

TABLE 2 Primary AI-based developmental-screening tools included in the scoping review, grouped by modality (16 studies, eight countries).

Tool	Modality	Country	Age (mo)	N	Sens (%)	Spec (%)	Status / AUC
EarlIPoint (Jones et al., 2023)	Eye-tracking, social-scene video	USA	16–30	disc./repl.	81.9/80.6	89.9/82.3	FDA 510(k) 2023
Gazefinder (Chetcuti et al., 2024)	Eye-tracking, 2-min protocol	Australia	9–14	54	feas. ^a	—	Commercial (JP)
Gaze metrics (Wang et al., 2024)	Eye-tracking, AVC/FAS/ASC	USA	18–84	39	88–100	80–88	Research (prelim.)
Canvas Dx (Megeerian et al., 2022)	Multimodal: caregiver Q + home video + clinician	USA	18–72	—	98.4	78.9	FDA De Novo 2021
Video–audio ens. (Natraj et al., 2024)	Multimodal video + audio NN	Switzerland	12–60	160	86.1 ^b	87.5	Research
Joint-Attention DL (Ko et al., 2023)	Computer-vision, joint-attention video	South Korea	24–72	95	99.2	95.5	AUC 0.996
Home-video ID (Kim et al., 2025)	Computer-vision, home video	South Korea	18–48	510	80.0	—	AUC 0.83
Stereotypy det. (Barami et al., 2024)	Computer-vision, movement video	Israel	<72	241	92.5	—	Research
SenseToKnow (Perochon et al., 2023; Aikat et al., 2025)	Tablet digital phenotyping	USA	17–36 ^c	475	87.8	80.8	AUC 0.90
ASDetect (Barbaro and Yaari, 2020)	Smartphone, parent-admin.	Australia	12–24	—	~83	NR	Research
Voice-acoustic (Briend et al., 2023)	Voice / acoustic ML	France	Children	108	91 ^d	—	Research
Well-baby EHR (Ben-Sasson et al., 2024)	EHR / ML risk prediction	Israel	0–24 ^e	780,610	75.0	81.0	AUC 0.86
Abbrev. M-CHAT-R (Pan et al., 2025)	Questionnaire-based ML	China	Toddlers	6,049	—	—	Research
CHAT-23 LightGBM (Lu et al., 2024)	Questionnaire-based ML	China	18–24	371	90.9	92.2	Research
Vietnamese M-CHAT-R (Van Vo et al., 2026)	Questionnaire-based ML	Vietnam	toddlers	—	—	—	Research
Movement sensors (Wilson et al., 2024)	Wearable sensors	USA	Infants ^f	—	—	—	Research

Metrics are as reported in the source studies and are not directly comparable across heterogeneous designs.

^aFeasibility/concurrent predictive validity, no fixed sens-spec.

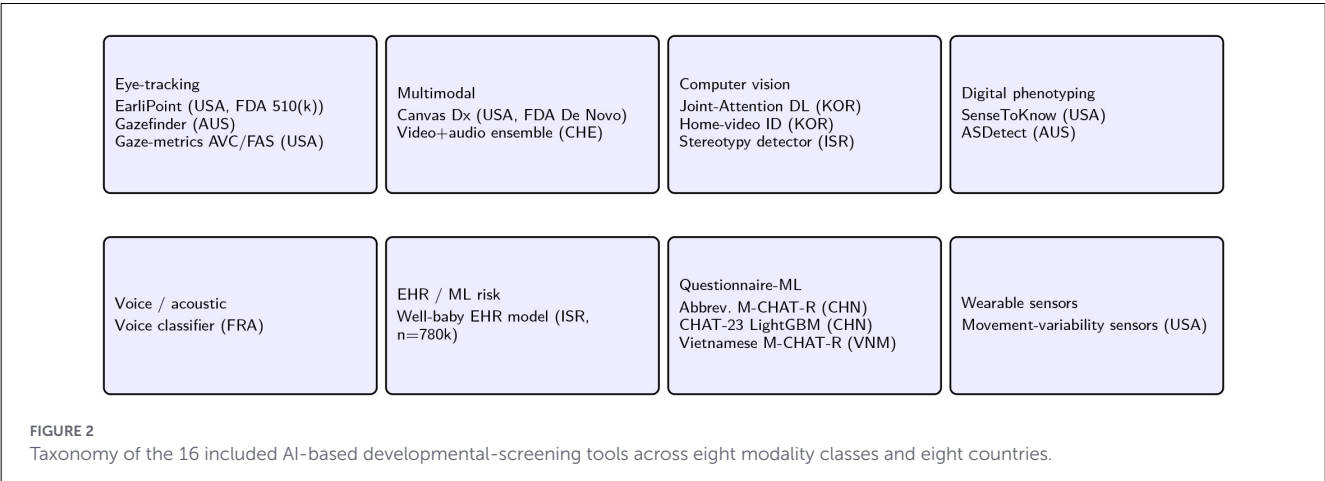
^bRecall.

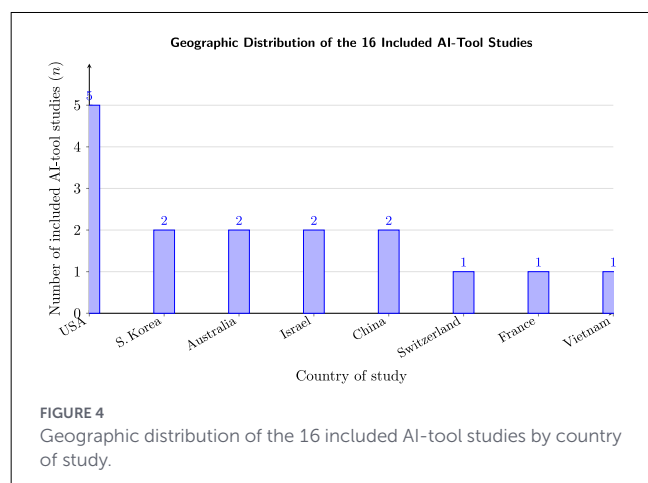
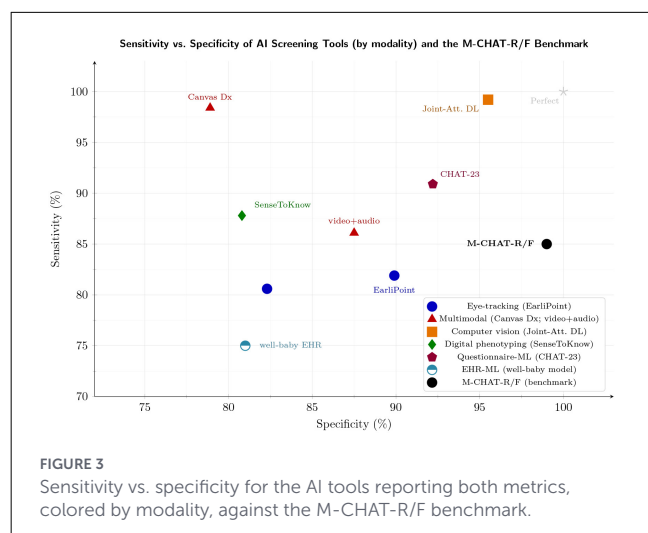
^cExtended to ages 3–8 in replication (Aikat et al., 2025).

^dVs. typically-developing children.

^ePredictors birth–24 month; mean diagnosis ~37.5 mo.

^fHigh-likelihood infant cohort. NR, not reported.





3.4 Regulatory trajectory and deployment readiness

The regulatory and publication trajectory is summarized in [Figure 5](#). Two FDA authorizations anchor the period: Canvas Dx (De Novo, 2021) and EarliPoint (510(k), 2023), both under software-as-a-medical-device (SaMD) pathways increasingly shaped by the FDA AI/ML action plan and pre-determined change-control guidance ([Singh et al., 2025](#)). In the European Union, AI medical devices fall under the Medical Device Regulation and the emerging AI Act ([Balogun et al., 2024](#)), yet no CE-marked autism-specific AI screening device was identified in the peer-reviewed literature, a notable transatlantic asymmetry. Beyond regulatory status, deployment readiness varies widely: EarliPoint requires specialty eye-tracking hardware; Canvas Dx yields a determinate output in roughly one third of referred cases; tablet and smartphone tools (SenseToKnow, ASDetect) are the most scalable but report the most modest PPV under community prevalence. Most research-stage tools have not undergone external or prospective validation, echoing broader pediatric-AI findings that few models reach routine workflows ([Wang et al., 2025](#); [Katonai et al., 2025](#)).

3.5 How this review differs from recent syntheses

[Table 1](#) positions our review against ten recent syntheses. Prior reviews are either modality-bounded, method-aggregating, technology-broad, or generative-AI-specific; none combines deployed-tool case studies with a structured equity matrix, a deployment-readiness lens, and a low-resource guidance-center implementation pathway.

4 Discussion

4.1 Summary and comparison with the traditional workflow

This review maps a field that has broadened substantially in modality and geography, from five tools across five modalities in early mappings to sixteen tools across eight modalities and eight countries here, yet remains concentrated on ASD and constrained by single-site, case-control validation. AI tools can equal or exceed questionnaire screeners on sensitivity, and sometimes objectivity and scalability, but the binding constraints are generalizability, equity, and operational integration rather than peak accuracy. No tool yet replaces universal M-CHAT-R/F at the 18–24-month visits; the realistic near-term role is a second-line, tiered aid consistent with American Academy of Pediatrics surveillance ([Hyman et al., 2020](#)).

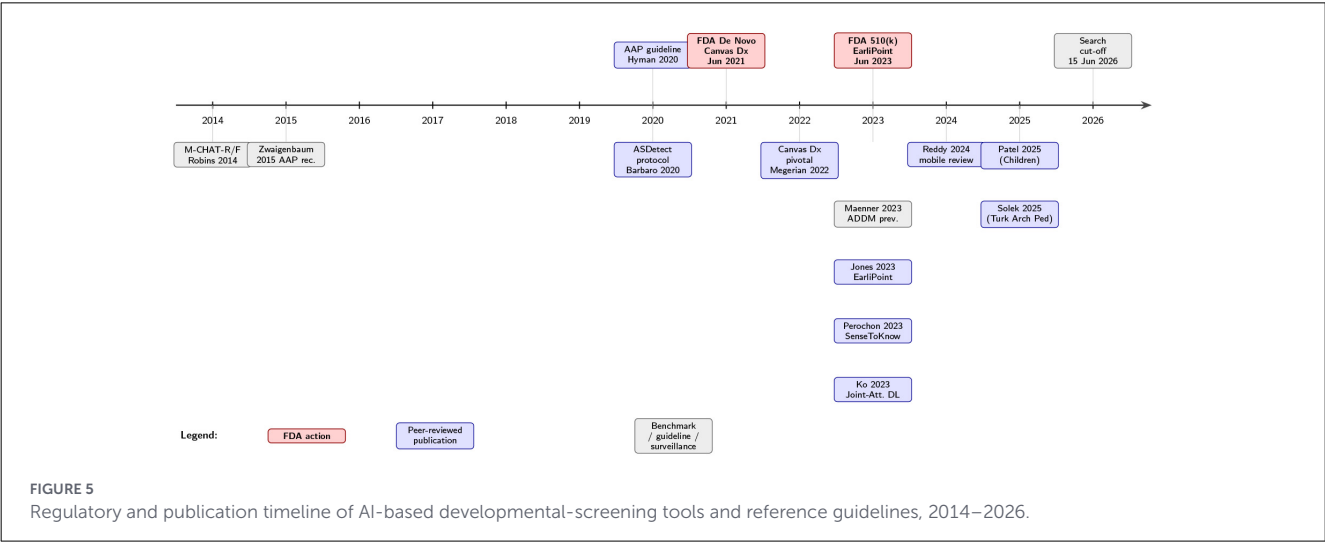
4.2 Methodological quality and risk of bias in the included evidence

Consistent with PRISMA-ScR and Joanna Briggs Institute guidance, scoping reviews map the extent and nature of evidence and do not require formal critical appraisal, and we therefore did not score the included studies with an instrument such as QUADAS-2. Nonetheless, charting the corpus made several recurring methodological limitations visible, and these should temper any reading of the accuracy figures reported above. First, and most consequentially, most tools were validated in relatively small, single-site, case-control samples that contrast already-diagnosed autistic children with typically-developing controls; such designs are well recognized to overstate accuracy relative to the unselected populations in which screening actually occurs (spectrum bias). Second, external and prospective validation remains rare: only SenseToKnow ([Perochon et al., 2023](#)) was evaluated prospectively in routine well-child visits, and only EarliPoint ([Jones et al., 2023](#)) reported an independent replication cohort, while most research-stage tools report internal cross-validation only, leaving overfitting and site-specific tuning unquantified. Third, reference standards are heterogeneous, ranging from full multidisciplinary diagnostic assessment to registry- or record-derived diagnosis, which shifts both the target condition and the apparent error rates. Fourth, the two FDA-authorized devices were evaluated in industry-sponsored pivotal trials in which the developer had a financial interest—an established

TABLE 3 Demographic equity and validation reporting across included tools (✓, reported/stratified; P, partial or noted without stratified metrics; —, not reported).

Tool	Sex	Race/ ethnicity	SES/language	External / prospective validation
SenseToKnow (Perochon et al., 2023)	✓	✓	—	✓ (prospective, well-child visits)
Well-baby EHR (Ben-Sasson et al., 2024)	✓	P	—	✓ (population-scale cohort)
EarliPoint (Jones et al., 2023)	—	P	—	✓ (independent replication cohort)
Canvas Dx (Megerian et al., 2022)	—	P	—	P (pivotal trial)
Home-video ID (Kim et al., 2025)	P ^a	—	—	Multi-site (nine hospitals)
Joint-Attention DL (Ko et al., 2023)	—	—	—	— (single center, <i>n</i> = 95)
Voice-acoustic (Briend et al., 2023)	—	—	—	— (single sample)
Questionnaire-ML (Pan et al., 2025; Lu et al., 2024; Van Vo et al., 2026)	—	—	—	Internal cross-validation only

Only one of 16 tools reported fully stratified performance.
^aMale pre-dominance (>2:1) noted; female/older-child generalizability flagged as a limitation. Remaining tools (Gazefinder, gaze metrics, stereotypy detector, ASDetect, wearables) reported no demographic stratification.



source of optimism bias in diagnostic-accuracy research—and reporting of thresholds and subgroup outcomes was frequently selective. Finally, demographic subgroup reporting was near-absent (Table 3), so the generalizability of the reported metrics to the populations most affected by diagnostic disparities is largely unverified.

Taken together, these features suggest that the corpus, while promising, sits at a relatively early stage of diagnostic-accuracy evidence, and that the highest reported figures are the least likely to survive prospective community deployment. A formal appraisal—for example QUADAS-2 for the diagnostic-accuracy studies and PROBAST for the prediction models—applied within a future systematic review would be a valuable and, we would argue, now necessary next step for this field.

4.3 Equity as the priority gap

A central promise of AI screening is to reduce disparities by providing low-cost, objective tools independent of parent

literacy or specialty-clinic access. The evidence is only partly aligned with that promise: SenseToKnow is the strongest equity exemplar (Perochon et al., 2023), but most tools report no stratified accuracy. Because diagnostic disparities and female under-detection are well documented (Pham et al., 2022; Aylward et al., 2021; Cruz et al., 2024) and clinical-AI fairness reviews repeatedly flag missing subgroup validation (Liu et al., 2025; Huang et al., 2022), we argue that stratified-accuracy reporting (by sex, race, ethnicity, primary language, and socioeconomic status) should be a default requirement for AI-screening trials, alongside child-centered ethical safeguards (Chng et al., 2025).

4.4 Regulatory and data-governance considerations

The FDA SaMD framework now accommodates adaptive AI through pre-determined change-control plans (Singh et al., 2025), while the EU MDR/AI Act imposes high-risk-device obligations (Balogun et al., 2024). The absence of any CE-marked

autism screening device suggests regulatory and market pathways outside the United States remain immature. Pediatric deployment further requires strict data governance: in the United States the Children's Online Privacy Protection Act, in the EU the General Data Protection Regulation provisions for minors, and in Türkiye the Personal Data Protection Law (KVKK, No. 6698). Notably, KVKK contains no child-specific provision (parental-consent obligations derive from the Civil Code), a gap relevant to any RAM-based deployment and to broader recommendations for pediatric AI data use (Muralidharan et al., 2023).

4.5 Implementation and adoption

Analytic accuracy is necessary but insufficient for adoption. Barriers identified across health-AI implementation reviews include clinician training and AI literacy, workflow and EHR integration, reimbursement uncertainty, trust, and accountability (Hassan et al., 2024; Katonai et al., 2025), and adoption is best understood through implementation-science frameworks such as NASSS (Greenhalgh et al., 2017). A child-development monitoring review likewise found acceptance and outcome evidence to be thin (Reinhart et al., 2024). SenseToKnow's deployment during well-child visits (Perochon et al., 2023) is an instructive exemplar of feasibility, but real-world effectiveness studies remain scarce.

4.6 Türkiye/RAM context: a low-resource implementation pathway

Türkiye's child-development pipeline involves the Ministry of National Education's pre-primary curriculum, the Family-and-Social-Services-Ministry Guidance and Research Centers (RAM), and Health-Ministry well-child clinics. AI screening tools are not yet routinely deployed in this pipeline, and only one Turkish-led synthesis on AI for ASD was identified (Solek et al., 2025). RAM centers, publicly funded, nationally distributed, and already tasked with developmental assessment and educational placement, are a natural locus for a pragmatic, equity-oriented pilot: a tablet digital-phenotyping or questionnaire-ML tool, culturally and linguistically adapted to Turkish, with prospective stratified-accuracy reporting. Such a pilot would simultaneously address the three gaps this review identifies (deployment evidence, equity validation, and non-Western/low-resource generalizability) and serve as a transferable template for other middle-income guidance-center systems.

4.7 Limitations

Four limitations qualify this review. First, and most importantly for interpretation, the search was restricted to English-language peer-reviewed primary studies. This restriction carries a language bias whose direction is predictable rather than random: tools developed and validated in non-English-speaking settings—for example in China, Japan, Latin America,

and Türkiye—are disproportionately likely to be reported in national-language journals and in databases we did not search, and are therefore under-represented in our map. Because one of our central findings is the geographic concentration of the evidence base, this bias may partly manufacture the very pattern we report, and the true modality and geographic breadth of the field is likely wider than Figure 4 suggests. Two considerations partially mitigate, but do not remove, this concern: several included tools originate from non-English-speaking countries (South Korea, Switzerland, France, Israel, China, Vietnam) yet were published in English, and regulatory milestones are documented in English-language registries. Future reviews should nonetheless search national-language sources (e.g., CNKI, SciELO, J-STAGE, ULAKBİM TR Dizin) with native-language strings, and we would expect such a search to soften—though probably not overturn—our conclusion about geographic concentration.

Second, we did not perform formal critical appraisal, consistent with PRISMA-ScR; the methodological concerns we identified are therefore descriptive rather than scored (see Section 4.2). Third, all charted primary studies target ASD, so AI work on language, motor, and ADHD precursors falls outside the mapped scope. Fourth, the reported metrics derive from heterogeneous designs, reference standards, age bands, and base rates, and are not directly comparable across tools; the performance ranges should be read as landscape descriptors, not as pooled estimates or a ranking.

5 Conclusion

The 2020–2026 evidence base for AI-based developmental screening of young children has matured in modality breadth and geographic reach: two FDA-authorized devices and a growing set of research prototypes span eye-tracking, multimodal, computer vision, digital phenotyping, voice, EHR, questionnaire-ML, and wearable approaches across eight countries. Yet the field remains autism-concentrated, equity-blind, and weakly translated into routine primary care, with no CE-marked autism screener and near-absent subgroup-stratified validation. Clinicians, child-development practitioners, and policymakers should treat current tools as emerging second-line aids that complement, not replace, validated questionnaire screeners, while the research community prioritizes equity validation, deployment-readiness reporting, and pragmatic implementation studies in low-resource guidance-center settings such as Türkiye's RAM network.

Author contributions

AA: Conceptualization, Validation, Resources, Writing – review & editing, Writing – original draft, Methodology. FG: Writing – original draft, Writing – review & editing, Conceptualization, Software, Visualization, Resources, Validation, Formal analysis, Methodology, Data curation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This study was supported by the Recep Tayyip Erdoğan University Development Foundation (grant number: 020260070080459).

Acknowledgments

The authors thank the staff of the Artificial Intelligence and Internet of Things Research Laboratory at Recep Tayyip Erdoğan University.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. AI was used to assist with drafting and language editing of the manuscript text and to generate the LaTeX/TikZ code used to produce the figures. The tool used was Claude (Anthropic; model Claude Opus 4.8; [https://](https://www.anthropic.com)

www.anthropic.com). All AI-assisted content was checked by the author(s) for factual accuracy and screened for plagiarism, and every cited reference was independently verified against its DOI. The initial and final prompts are provided in the [Supplementary material](#). Generative AI is not listed as an author.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2026.1913332/full#supplementary-material>

DATA SHEET 1

Full database search strategies, PRISMA-ScR charting framework, and the initial and final generative-AI prompts.

References

- Agrawal, R., and Agrawal, R. (2025). Early autism detection: a review of emerging technologies, biomarkers, and explainable ai approaches. *Mol. Brain* 19:7. doi: 10.1186/s13041-025-01269-9
- Aikat, V., Carpenter, K. L. H., Babu, P. R. K., Di Martino, J. M., Espinosa, S., Compton, S., et al. (2025). Autism digital phenotyping in pre-school- and school-age children. *Autism Research* 18, 1217–1233. doi: 10.1002/aur.70032
- Aylward, B. S., Gal-Szabo, D. E., and Taraman, S. (2021). Racial, ethnic, and sociodemographic disparities in diagnosis of children with autism spectrum disorder. *J. Dev. Behav. Pediatrics* 42, 682–689. doi: 10.1097/DBP.0000000000000996
- Balogun, E., Dcosta, D., Boch, A., and Luetge, C. (2024). Exploring key stakeholders' perspectives on integrating the eu ai act with the mdr for certifying ai medical devices. *AI and Ethics* 5, 2999–3013. doi: 10.1007/s43681-024-00612-5
- Barami, T., Manelis-Baram, L., Kaiser, H., Ilan, M., Slobodkin, A., Hadashi, O., et al. (2024). Automated analysis of stereotypical movements in videos of children with autism spectrum disorder. *JAMA Netw. Open* 7:e2432851. doi: 10.1001/jamanetworkopen.2024.32851
- Barbaro, J., and Yaari, M. (2020). Study protocol for an evaluation of ASDetect—a mobile application for the early detection of autism. *BMC Pediatrics* 20:21. doi: 10.1186/s12887-019-1888-6
- Ben-Sasson, A., Guedalia, J., Nativ, L., Ilan, K., Shaham, M., and Gabis, L. V. (2024). A prediction model of autism spectrum diagnosis from well-baby electronic data using machine learning. *Children* 11:429. doi: 10.3390/children11040429
- Briend, F., David, C., Silleri, S., Malvy, J., Ferré, S., and Latinus, M. (2023). Voice acoustics allow classifying autism spectrum disorder with high accuracy. *Transl. Psychiatry* 13:250. doi: 10.1038/s41398-023-02554-8
- Chetcuti, L., Varcin, K. J., Boutrus, M., Smith, J., Bent, C. A., Whitehouse, A. J. O., et al. (2024). Feasibility of a 2-minute eye-tracking protocol to support the early identification of autism. *Sci. Rep.* 14:5117. doi: 10.1038/s41598-024-55643-z
- Chng, S. Y., Tern, M. J. W., Lee, Y. S., Cheng, L. T.-E., Kapur, J., Eriksson, J. G., et al. (2025). Ethical considerations in ai for child health and recommendations for child-centered medical ai. *NPJ Digital Medicine* 8:152. doi: 10.1038/s41746-025-01541-1
- Cruz, S., Zubizarreta, S. C.-P., Costa, A. D., Araújo, R., Martinho, J., Tubío-Funqueriño, M., et al. (2024). Is there a bias toward males in the diagnosis of autism? a systematic review and meta-analysis. *Neuropsychol. Rev.* 35, 153–176. doi: 10.1007/s11065-023-09630-2
- Greenhalgh, T., Wherton, J., Papoutsis, C., Lynch, J., Hughes, G., A'Court, C., et al. (2017). Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J. Med. Internet Res.* 19:e367. doi: 10.2196/jmir.8775
- Hasan, H. H. (2025). Artificial intelligence in early detection of autism spectrum disorder for preschool ages: a systematic literature review. *medRxiv* (Preprint). doi: 10.1101/2025.09.10.25335333
- Hassan, M., Kushniruk, A., and Borycki, E. (2024). Barriers to and facilitators of artificial intelligence adoption in health care: scoping review. *JMIR Human Factors* 11:e48633. doi: 10.2196/48633
- Huang, J., Galal, G., Etemadi, M., and Vaidyanathan, M. (2022). Evaluation and mitigation of racial bias in clinical machine learning models: scoping review. *JMIR Med. Informat.* 10:e36388. doi: 10.2196/36388
- Hyman, S. L., Levy, S. E., and Myers, S. M. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics* 145:e20193447. doi: 10.1542/9781610024716-part01-ch002
- Jones, W., Klaiman, C., Richardson, S., Lambha, M., Reid, M., Hamner, T., et al. (2023). Development and replication of objective measurements of social visual engagement

- to aid in early diagnosis and assessment of autism. *JAMA Netw. Open* 6:e2330145. doi: 10.1001/jamanetworkopen.2023.30145
- Katonai, G., Arvai, N., and Mesko, B. (2025). Ai and primary care: scoping review. *J. Med. Int. Res.* 27:e65950–e65950. doi: 10.2196/65950
- Kim, D. Y., Do, R., Shin, Y., Sim, H., Kim, H., Cho, S., et al. (2025). Automated AI based identification of autism spectrum disorder from home videos. *NPJ Digital Med.* 8:607. doi: 10.1038/s41746-025-01993-5
- Ko, C., Lim, J.-H., Hong, J., Hong, S.-B., and Park, Y. R. (2023). Development and validation of a joint attention-based deep learning system for detection and symptom severity assessment of autism spectrum disorder. *JAMA Netw. Open* 6:e2315174. doi: 10.1001/jamanetworkopen.2023.15174
- Liu, M., Ning, Y., Teixayavong, S., Liu, X., Mertens, M., Shang, Y., et al. (2025). A scoping review and evidence gap analysis of clinical ai fairness. *NPJ Digital Med.* 8:360. doi: 10.1038/s41746-025-01667-2
- Lu, H., Zhang, H., Zhong, Y., Meng, X.-Y., Zhang, M.-F., and Qiu, T. (2024). A machine learning model based on chat-23 for early screening of autism in chinese children. *Front. Pediatr.* 12:1400110. doi: 10.3389/fped.2024.1400110
- Maenner, M. J., Warren, Z., Williams, A. R., and others (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, united states, 2020. *MMWR Surveillance Summaries* 72:1–14. doi: 10.15585/mmwr.ss7202a1
- Megerian, J. T., Dey, S., Melmed, R. D., Coury, D. L., Lerner, M. A., Nicholls, C. J., et al. (2022). Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder. *NPJ Digital Med.* 5:57. doi: 10.1038/s41746-022-00598-6
- Mohammadi, F., Shahrokhi, H., Asadzadeh, A., Pirmoradi, S., Moghtader, A., and Rezaei-Hachesu, P. (2025). Artificial intelligence in autism spectrum disorder diagnosis: a scoping review of face, voice, and text analysis methods. *Health Sci. Rep.* 8:e71476. doi: 10.1002/hsr.2.71476
- Muralidharan, V., Burgart, A., Daneshjou, R., and Rose, S. (2023). Recommendations for the use of pediatric data in artificial intelligence and machine learning accept-AI. *NPJ Digital Med.* 6:166. doi: 10.1038/s41746-023-00898-5
- Natraj, S., Kojovic, N., Maillart, T., and Schaer, M. (2024). Video-audio neural network ensemble for comprehensive screening of autism spectrum disorder in young children. *PLoS ONE* 19:e0308388. doi: 10.1371/journal.pone.0308388
- Pan, N., Chen, L., Wu, B., Chen, F., Chen, J., Huang, S., et al. (2025). Developing a simplified measure to predict the risk of autism spectrum disorders: abbreviating the m-chat-r using a machine learning approach in china. *Psychiatry Research* 344:116353. doi: 10.1016/j.psychres.2025.116353
- Patel, R., and others (2025). AI-enabled technologies and biomarker analysis for the early identification of autism and related neurodevelopmental disorders. *Children* 12:1670. doi: 10.3390/children12121670
- Perochon, S., Di Martino, J. M., Carpenter, K. L. H., Compton, S., Davis, N., Eichner, B., et al. (2023). Early detection of autism using digital behavioral phenotyping. *Nature Med.* 29, 2489–2497. doi: 10.1038/s41591-023-02574-3
- Pham, H. H., Sandberg, N., Trinkl, J., and Thayer, J. (2022). Racial and ethnic differences in rates and age of diagnosis of autism spectrum disorder. *JAMA Netw. Open* 5:e2239604. doi: 10.1001/jamanetworkopen.2022.39604
- Rakotomanana, H., and Rouhafzay, G. (2025). A scoping review of ai-based approaches for detecting autism traits using voice and behavioral data. *Bioengineering* 12:1136. doi: 10.3390/bioengineering12111136
- Reddy, K., Taksande, A., and Kurian, B. (2024). Harnessing the power of mobile phone technology: Screening and identifying autism spectrum disorder with smartphone apps. *Cureus* 16:e55004. doi: 10.7759/cureus.55004
- Reinhart, L., Bischops, A. C., Kerth, J.-L., Hagemester, M., Heinrichs, B., Eickhoff, S. B., et al. (2024). Artificial intelligence in child development monitoring: A systematic review on usage, outcomes and acceptance. *Intelligence-Based Med.* 9:100134. doi: 10.1016/j.ibmed.2024.100134
- Robins, D. L., Casagrande, K., Barton, M., Chen, C.-M. A., Dumont-Mathieu, T., and Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics* 133, 37–45. doi: 10.1542/peds.2013-1813
- Shen, Z., and Yu, C.-L. (2025). How technology advances research and practice in autism spectrum disorder: A narrative review on early detection, subtype stratification, and intervention. *Brain Sci.* 15:890. doi: 10.3390/brainsci15080890
- Singh, V., Cheng, S., Kwan, A. C., and Ebinger, J. (2025). United states food and drug administration regulation of clinical software in the era of artificial intelligence and machine learning. *Mayo Clinic Proceedings: Digital Health* 3:100231. doi: 10.1016/j.mcpdig.2025.100231
- Sohn, J.-S., Lee, E., Kim, J.-J., Oh, H.-K., and Kim, E. (2025). Implementation of generative ai for the assessment and treatment of autism spectrum disorders: a scoping review. *Front. Psychiatry* 16:1628216. doi: 10.3389/fpsy.2025.1628216
- Solek, P., Nurfutri, E., Sahril, I., Prasetya, T., and Rizqiamuti, A. F., Burhan, et al. (2025). The role of artificial intelligence for early diagnostic tools of autism spectrum disorder: a systematic review. *Turkish Archiv. Pediatrics* 60, 126–140. doi: 10.5152/TurkArchPediatr.2025.24183
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Internal Med.* 169, 467–473. doi: 10.7326/M18-0850
- Van Vo, T., Nguyen, P. M., Nguyen, D. N., Van Nguyen, T., Thai, D. M., Ly, H. H. V., et al. (2026). Machine learning-assisted autism risk stratification in toddlers using the vietnamese m-chat-r and perinatal predictors: a cross-sectional study in Vietnam. *J. Autism Dev. Disord.* doi: 10.1007/s10803-026-07227-1. [Epub ahead of print].
- Wang, Q., Yin, J., Zhang, X., Ou, H., Li, F., Zhang, Y., et al. (2025). Applications of artificial intelligence in early childhood health management: a systematic review from fetal to pediatric periods. *Front. Pediatrics* 13. doi: 10.3389/fped.2025.1613150
- Wang, R. K., Kwong, K., Liu, K., and Kong, X.-J. (2024). New eye tracking metrics system: the value in early diagnosis of autism spectrum disorder. *Frontiers in Psychiatry* 15. doi: 10.3389/fpsy.2024.1518180
- Wilson, R. B., Vangala, S., Reetzke, R., Piergies, A., Ozonoff, S., and Miller, M. (2024). Objective measurement of movement variability using wearable sensors predicts <sc>asd</sc> outcomes in infants at high likelihood for <sc>asd</sc> and <sc>adhd</sc>. *Autism Res.* 17, 1094–1105. doi: 10.1002/aur.3150
- Zhang, S. (2025). Ai-assisted early screening, diagnosis, and intervention for autism in young children. *Front. Psychiatry* 16:1513809. doi: 10.3389/fpsy.2025.1513809
- Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., et al. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics* 136, S41–S59. doi: 10.1542/peds.2014-3667D