Vol. 8, No. 4, August 2026

Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Toward AI-Based Child Emotion Early Warning Systems in Child-Oriented Digital Environments: A Socio-Technical Systematic Review

Ratih Titi Komala Sari[1*], Widowati [2], Adi Wibowo [3]

[1]Doctoral Program of Information Systems, Postgraduate School, Diponegoro University, Indonesia

[2]Mathematics Department, Faculty of Mathematics and Sciences, Diponegoro University, Indonesia

[3]Informatics Department, Faculty of Mathematics and Sciences, Diponegoro University, Semarang, Indonesia

Received:
February 1, 2026
Revised:
July 19, 2026
Accepted:
September 2, 2026
Published:
September 3, 2026

Corresponding Author:
Author Name*:
Ratih Titi Komala Sari
Email*:
ratihtiti@students.undip.ac.id

DOI:
10.63158/journalisi.v8i4.1811

Abstract. AI-based emotion recognition is gaining increasing attention in child-oriented digital environments, including digital play, AI toys, online learning, therapeutic platforms, and interactive child-computer systems. However, the translation of recognized emotions into responsible caregiver-facing early warnings remains insufficiently explored. This systematic literature review synthesizes technical and socio-technical perspectives on child emotion recognition, focusing on sensing modalities, AI approaches, datasets, alert interpretation, trust, privacy, ethics, governance, and caregiver acceptance. Following PRISMA 2020 guidelines, 400 records from Scopus and IEEE Xplore were screened, resulting in 32 studies included in the core synthesis. The findings reveal that facial-expression analysis and convolutional neural networks dominate current research, while child-specific datasets, multimodal learning, real-world validation, uncertainty communication, privacy-by-design, and caregiver-centered evaluation remain limited. This review proposes a socio-technical framework linking AI emotion inference with explainable alert translation, privacy-aware governance, and caregiver decision support. The proposed emotion-to-alert mechanism remains conceptual and requires empirical validation before practical deployment.

Keywords: child emotion recognition; artificial intelligence; caregiver decision support; explainable AI; privacy-aware systems

1. INTRODUCTION

AI-driven emotion recognition is an established research stream in affective computing, human-computer interaction, educational technology, and child-oriented digital systems. Previous reviews show that emotion recognition has advanced through facial-expression analysis, speech processing, physiological sensing, and multimodal fusion, supported by machine learning and deep learning algorithms [1]-[5]. These advances create opportunities for systems that do not merely classify emotions but also support adaptive responses in digital interaction environments.

For children, emotion recognition requires more cautious interpretation than adult-centered affective computing. Children may express emotions through subtle, spontaneous, and developmentally specific facial, vocal, and behavioral cues. Reviews of technology-assisted emotion recognition for autistic children show that facial expression, speech prosody, and physiological signals are frequently used; however, challenges remain in stimulus design, labeling procedures, multimodal fusion, and open child-specific datasets [6], [7]. These limitations are relevant to child-oriented digital environments because many systems still rely on adult or mixed-age datasets.

In this review, child-oriented digital environments refer to digitally mediated settings designed for, adapted to, or substantially involving children, including digital play, serious games, AI toys, AR/VR activities, online learning platforms, digital therapeutic applications, and interactive child-computer systems. Digital play is treated as a central but not exclusive context, while online learning and therapeutic applications are included only when they involve child-oriented interaction and emotion-related AI, rather than general classroom monitoring or clinical diagnosis alone. Therapeutic applications were included when they functioned as digitally mediated interactive environments in which emotion-related information influenced feedback, adaptation, interaction, or caregiver-relevant interpretation, rather than because of their clinical diagnostic purpose.

In this review, the term child is used broadly to include early childhood, preschool, school-age children, and youth when the study context involves child-oriented digital interaction, emotion-related AI, or caregiver-relevant affective information. This broad scope is necessary because the reviewed literature does not consistently use a single age

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

boundary. Accordingly, no single numerical age threshold was imposed across the review; reported participant ages were extracted at study level where available, and findings were interpreted according to the developmental categories reported by the original studies. However, the interpretation of emotional signals remains developmentally sensitive. Crying, fear, sadness, anger, frustration, and withdrawal may have different meanings depending on age, developmental stage, communication ability, play context, and caregiver knowledge of the child.

Studies involving autistic children or neurodevelopmental conditions were treated as eligible when they addressed child-oriented digital interaction, emotion recognition, affective support, or caregiver-relevant emotion interpretation. These studies were not treated as universally generalizable evidence for all children. Instead, they were interpreted as an important subgroup that highlights population-specific challenges, including atypical emotional expression, differences in social communication, dataset representativeness, annotation reliability, and the need for cautious, context-sensitive alert translation.

This review distinguishes three related concepts. Emotion recognition refers to the computational inference of emotional states from facial, vocal, behavioral, physiological, or multimodal signals. Emotion-responsive systems use such inference to adapt feedback, interaction, learning support, or therapeutic activities. Early warning systems extend this by translating uncertain emotion predictions into understandable, proportionate, and caregiver-facing alerts. This distinction is important because most studies focus on recognition performance or adaptive interaction, while fewer address alert escalation, caregiver notification, and follow-up action.

The primary users of early warning systems are parents and caregivers, who are responsible for interpreting and responding to children's emotional distress outside clinical settings. Teachers, therapists, and platform providers are considered supporting stakeholders influencing system design, deployment, governance, and ethics. From an Information Systems perspective, the central contribution of this review is not limited to AI-based emotion classification. The review focuses on the socio-technical layer through which predicted emotional states are translated into explainable, proportionate, and privacy-sensitive alerts for caregivers. This includes alert design, dashboard-based decision support, uncertainty communication, privacy governance, caregiver trust, and responsible acceptance of emotional AI in child-oriented digital environments.

Existing reviews have explored emotion recognition techniques, multimodal affective computing, learning environments, and ethical concerns, but rarely integrate these from a socio-technical perspective linking sensing modalities, AI methods, child-specific datasets, caregiver alerts, privacy, trust, and acceptance. Therefore, this review synthesises literature on AI-driven emotion recognition and emotion-responsive systems involving children, contributes a socio-technical framework toward AI-based child emotion early warning systems, and identifies directions for future empirical validation.

2. METHODS

2.1. PRISMA Review Design

This study used a PRISMA 2020-guided systematic literature review [8], [9] with a taxonomy-based narrative synthesis due to heterogeneous literature in design, data, AI methods, and contexts. The review synthesised both technical evidence on child emotion recognition and socio-technical aspects such as caregiver alerts, trust, privacy, ethics, governance, and acceptance.

The study separated the core PRISMA corpus from a supplemental source audit. The core corpus supported the PRISMA diagram, RQ mapping, and frequency analysis, while the supplemental audit only documented source provenance and potential additional studies. This ensured that only fully assessed studies contributed to the synthesis. Five research questions were addressed, covering sensing modalities (RQ1), AI methods (RQ2), datasets and populations (RQ3), caregiver alert translation (RQ4), and socio-technical factors (RQ5).

Scopus and IEEE Xplore were designated as the core databases because the formal PRISMA screening workflow was initially executed on these two sources. Scopus provided multidisciplinary coverage across computer science, information systems, education, psychology, and technology-related research, whereas IEEE Xplore complemented the search with engineering-oriented literature in artificial intelligence, affective computing, sensing, and interactive systems. SpringerLink, PubMed, ScienceDirect, and Emerald were subsequently examined as supplemental sources to assess source provenance and potential evidence expansion. Because these additional records had not completed the

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

same harmonized eligibility assessment, quality appraisal, and structured extraction workflow for all candidates, they were not retrospectively merged into the fixed core corpus.

2.2. Search Strategy and Database Fields

Core data were retrieved from Scopus and IEEE Xplore on 17 June 2026. Scopus was chosen for its multidisciplinary coverage, while IEEE Xplore was selected for its relevance to AI and sensing technologies. Searches used TITLE-ABS-KEY in Scopus and metadata fields (title, abstract, index terms) in IEEE Xplore, limited to English peer-reviewed publications from 2020–2026.

The search combined five blocks: emotion recognition, child population, digital environments, AI methods, and socio-technical factors. The mandatory socio-technical search block was intended to align retrieval with the Information Systems focus of this review, particularly trust, privacy, ethics, acceptance, perceived risk, governance, and caregiver relevance. However, this choice introduced a sensitivity–specificity trade-off: technically relevant child emotion-recognition studies may have been missed when they did not explicitly mention socio-technical terms in searchable metadata. Future updates should therefore use two complementary search streams: a technical search focused on child emotion-recognition methods and a socio-technical search focused on governance, trust, privacy, ethics, acceptance, caregiver alerts, and responsible implementation. The two evidence streams can subsequently be integrated during full-text screening and synthesis.

1) Scopus search string:

TITLE-ABS-KEY (("emotion recognition" OR "affective computing" OR "emotional AI" OR "automated emotion recognition" OR "facial emotion recognition")
AND (child* OR children OR toddler* OR preschool OR "early childhood")
AND ("digital play" OR game* OR "serious game*" OR "online learning" OR "interactive system" OR "AI toy*" OR "child-computer interaction")
AND ("deep learning" OR "machine learning" OR CNN OR "convolutional neural network" OR "vision transformer" OR ViT OR multimodal OR "facial expression recognition")
AND (trust OR privacy OR ethics OR acceptance OR "parental acceptance" OR "technology acceptance" OR "perceived risk"))

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2) IEEE Xplore search string:

("emotion recognition" OR "affective computing" OR "emotional AI" OR "automated emotion recognition" OR "facial emotion recognition")
AND (child* OR children OR toddler* OR preschool OR "early childhood")
AND ("digital play" OR game* OR "serious game*" OR "online learning" OR "interactive system" OR "AI toy*" OR "child-computer interaction")
AND ("deep learning" OR "machine learning" OR CNN OR "convolutional neural network" OR "vision transformer" OR ViT OR multimodal OR "facial expression recognition")
AND (trust OR privacy OR ethics OR acceptance OR "parental acceptance" OR "technology acceptance" OR "perceived risk")

For the supplemental source audit, the master search logic was adapted to the syntax and metadata structure supported by SpringerLink, PubMed, ScienceDirect, and Emerald. The conceptual search blocks—emotion recognition, child population, child-oriented digital context, AI or computational method, and socio-technical factors—were preserved where supported by the database interface. Exact syntax, searchable fields, phrase-search behavior, and export capabilities differed across platforms; therefore, the supplemental searches should be interpreted as adapted source-provenance and sensitivity searches rather than exact replications of the core Scopus and IEEE Xplore search. The low IEEE Xplore yield is consistent with the restrictive conjunction of the five mandatory search blocks. However, because no search-block ablation or sensitivity analysis was performed, the low yield cannot be attributed exclusively to the socio-technical block. Future sensitivity searches should compare retrieval with and without mandatory socio-technical terms.

2.3. Eligibility Criteria

Studies were included if they addressed AI-based emotion recognition involving children or child-oriented digital systems such as games, AI toys, AR/VR, online learning, or interactive platforms. Exclusions covered adult-only studies, unrelated affective computing, non-interactive clinical studies, and non-peer-reviewed or inaccessible sources. Autism and neurodevelopmental studies were included when relevant but treated as specific contexts. Mixed-population studies were included only if child-related contributions were clear.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 1. Review protocol summary

Component	Specification
Review type	Systematic literature review with taxonomy-based narrative synthesis
Guideline	PRISMA 2020
Core databases	Scopus and IEEE Xplore
Supplemental source audit	SpringerLink, PubMed, ScienceDirect, and Emerald; used only for source-provenance checking and contextual interpretation
Publication year	2020-2026
Document type	Peer-reviewed journal articles and reputable conference papers
Core domains	AI-based emotion recognition, child emotion recognition, child-oriented digital environments including digital play, caregiver-facing alerts, trust, privacy, ethics, governance, and caregiver or parental acceptance
Synthesis method	Descriptive mapping, thematic synthesis, gap analysis, and conceptual framework development
Search and export dates	Scopus and IEEE Xplore: 17 June 2026. File-stamped supplemental exports: SpringerLink, 3 May 2026; PubMed, 11 April 2026; ScienceDirect and Emerald, 2 June 2026.
Evidence status	The PRISMA flow diagram, Figures 2–6, and all RQ-specific frequencies are derived exclusively from the core full-text screened corpus (n = 32). Supplemental records are reported only as source-provenance and contextual audit evidence because 219 DOI-level candidates remain pending equivalent review.
Supplemental audit boundary	Details of the supplemental source audit are reported in Appendix A and are not included in the core PRISMA counts.

Table 2. Eligibility criteria

Criterion	Inclusion criteria	Exclusion criteria
Topic	Studies discussing emotion recognition, affective	Studies unrelated to emotion recognition or affective computing

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Criterion	Inclusion criteria	Exclusion criteria
	computing, emotional AI, or automated emotion recognition	
Population/context	Studies involving children, early childhood, preschool, youth, parents, caregivers, or child-oriented systems	Studies focused only on adults without child-related implications
Technology	Studies using AI, machine learning, deep learning, computer vision, speech processing, sensors, or multimodal fusion	Studies without AI, sensing, or computational analysis
Application	Child-oriented digital environments, including digital play, games, serious games, AI toys, AR/VR, online learning, digital therapeutic applications, interactive systems, and child-computer interaction, when emotion recognition or emotion-responsive interaction is central to the study	Generic emotion studies without a child-oriented digital context; general classroom monitoring without emotion-recognition relevance; adult-only systems without child-related implications; clinical diagnosis studies without interactive or information-system relevance
Socio-technical factor	Studies addressing trust, privacy, ethics, governance, caregiver or parental acceptance, or perceived risk	Studies lacking relevance to the research questions after full-text screening
Publication	English peer-reviewed sources from 2020–2026	Non-peer-reviewed sources, inaccessible full texts, duplicates, or editorials without method

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

2.4. Screening, Quality Appraisal, and Consensus Procedure

The process included identification, screening, eligibility, quality appraisal, and inclusion. From 400 records (398 Scopus, 2 IEEE), 69 were assessed in full text, and 32 were included after exclusions. Screening and extraction were conducted by the first author, with decisions validated through co-author consensus. Because the workflow used single-reviewer screening and initial extraction followed by co-author validation rather than independent dual screening, no formal inter-rater agreement statistic was calculated. Uncertain classifications and disagreements identified during co-author validation were discussed until a consensus decision was reached. Quality appraisal used seven criteria: objective clarity, relevance to child or child-oriented digital-environment context, dataset or data-source transparency, AI method or system clarity, evaluation adequacy, attention to ethics/privacy/bias, and limitations disclosure. Each criterion was scored from 0 to 2, yielding a maximum score of 14. Scores of 11–14 were classified as high quality, 7–10 as moderate quality, and 0–6 as low quality. The study-level quality-appraisal matrix, including C1–C7 scores, total scores, quality categories, and key methodological weaknesses, is provided in Supplementary Table S2.

2.5. Data Extraction and Non-Exclusive Coding

Data extraction captured study details, population, modality, dataset, AI method, system function, evaluation, and socio-technical aspects. Coding was non-exclusive, allowing studies to belong to multiple categories. The data-extraction form included bibliographic information, research context, child age group, population type, sensing modality, dataset, AI method, emotion labels, system function, evaluation metrics, stakeholder involvement, trust-related factors, privacy and ethical issues, key findings, limitations, and relevance to caregiver-facing early warning systems. The structured data-extraction form is provided in Supplementary File S3, and the seven-criterion quality-appraisal codebook, including scoring definitions and quality-category thresholds, is provided in Supplementary File S4.

A structured coding guide was used to operationalize category assignment across RQ1–RQ5. Coding was non-exclusive: a study could receive more than one code within a research question when multiple modalities, methods, populations, system functions, or socio-technical factors were explicitly reported. For example, a study involving autistic children and facial-expression recognition could simultaneously contribute to the child

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

population, ASD/neurodevelopmental subgroup, facial/visual modality, AI-method, dataset, and ethics/privacy categories where applicable. Studies involving multiple populations were coded into each applicable population category; therefore, category frequencies are non-exclusive and should not be summed to derive the total number of studies Analysis used descriptive mapping, thematic synthesis, and gap analysis. RQ1–RQ2 focused on technical aspects, RQ3 on datasets and populations, RQ4 on caregiver alerts, and RQ5 on socio-technical factors.

2.6. Supplemental Source Audit

A supplemental audit from SpringerLink, PubMed, ScienceDirect, and Emerald identified additional studies and verified source provenance. Duplicate detection used DOI or title matching. These records were excluded from the core synthesis because 219 candidates had not completed full evaluation. Consequently, supplemental records were not used to calculate PRISMA counts, RQ-specific frequencies, evidence-map frequencies, or core quality-appraisal statistics. They are reported separately in Appendix A for source-provenance documentation and future review expansion.

2.7. Supplementary Materials and Reproducibility

To strengthen methodological transparency, Supplementary Table S1 provides the study-level characteristics of the 32 core included studies, Supplementary Table S2 provides study-level C1–C7 quality-appraisal scores, Supplementary File S3 provides the structured data-extraction form, and Supplementary File S4 provides the seven-criterion quality-appraisal codebook. A study-level list of full-text exclusions and reasons is provided in Supplementary Table S5. Supplemental source-provenance and staged-screening information is reported separately in Appendix A.

3. RESULTS AND DISCUSSION

The core database searches yielded 400 records: 398 from Scopus and two from IEEE Xplore. No duplicate records were identified during core deduplication. Following title-and-abstract screening, 69 reports underwent full-text eligibility assessment. Thirty-seven reports were excluded according to the predefined criteria, leaving 32 studies for qualitative synthesis and RQ-based evidence mapping.

A supplemental audit was conducted to document source provenance beyond Scopus and IEEE Xplore. Because 219 DOI-level candidates had not completed equivalent full-text eligibility review, quality appraisal, and structured data extraction, these records were not incorporated into the RQ-specific synthesis, PRISMA count, or evidence-map frequencies. Details of the supplemental audit are provided in Appendix A.

The RQ-specific graphs are based exclusively on the core included corpus (n = 32). Counts are non-exclusive because a study may address more than one modality, method family, dataset context, or socio-technical factor. The graphs should therefore be interpreted as evidence maps rather than mutually exclusive frequency distributions. To improve transparency while preserving readability, Table 3 presents aggregate characteristics of the core corpus, while complete study-level characteristics are reported in Supplementary Table S1.

Table 3. Aggregate characteristics of the 32 core included studies

Evidence component	**Summary**
Core included studies	32 studies
Main evidence types	Technical emotion-recognition studies, systematic reviews, child-oriented systems, ethical/governance studies
Dominant modalities	Facial/visual, physiological/sensor, speech/audio, multimodal sensing
Dominant AI/system families	CNN/deep CNN, multimodal fusion, Transformer/ViT, classical ML, XAI, real-time/IoT systems
Main population contexts	Children/young children, ASD/ADHD contexts, youth/students, parents/caregivers, mixed-age datasets
Supplementary extraction table	Full study-level extraction is provided in Supplementary Table S1

Supplementary Table S1, provided as an accompanying supplementary file, contains the complete Core_Studies_32 extraction and quality-appraisal matrix for the 32 core included studies. The quality-appraisal summary is presented in Table 4. Each study was assessed using seven criteria with a maximum score of 14. Studies scoring 11–14 were classified as high quality, 7–10 as moderate quality, and 0–6 as low quality. The most frequent weaknesses involved limited child-specific dataset transparency, incomplete

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

reporting of ethics/privacy considerations, limited ecological validation, and brief disclosure of study limitations.

Table 4. Quality-appraisal summary of the 32 core studies

Metric	Value	Interpretation
Core included studies scored	32	Core corpus remains fixed.
High quality count	28	High = 11–14
Moderate quality count	4	Moderate = 7–10
Low quality count	0	Low = 0–6
Average QA score	12.03	Mean of C1–C7 total score
Minimum QA score	9	Lowest score among S1–S32
Maximum QA score	14	Highest score among S1–S32

Note. The complete study-level quality-appraisal matrix, including C1–C7 scores, total score, quality category, and key methodological weakness for each core study, is provided in Supplementary Table S2.

3.1. RQ1. Sensing modalities for child emotion recognition

Figure 2 shows that facial or visual sensing was the most frequently represented modality, appearing in 19 of the 32 included studies. Physiological or sensor-based signals and speech or audio was each coded in seven studies, followed by multimodal or multichannel sensing in six studies and gameplay or interaction traces in five studies. Text or dialogue and gesture or pose appeared in two studies each, while eye tracking or gaze appeared in one study. Because the categories are non-exclusive, several studies combined more than one sensing channel. The pattern indicates that camera-based facial analysis remains the main technical foundation of the corpus, largely because it can be integrated into existing digital environments.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

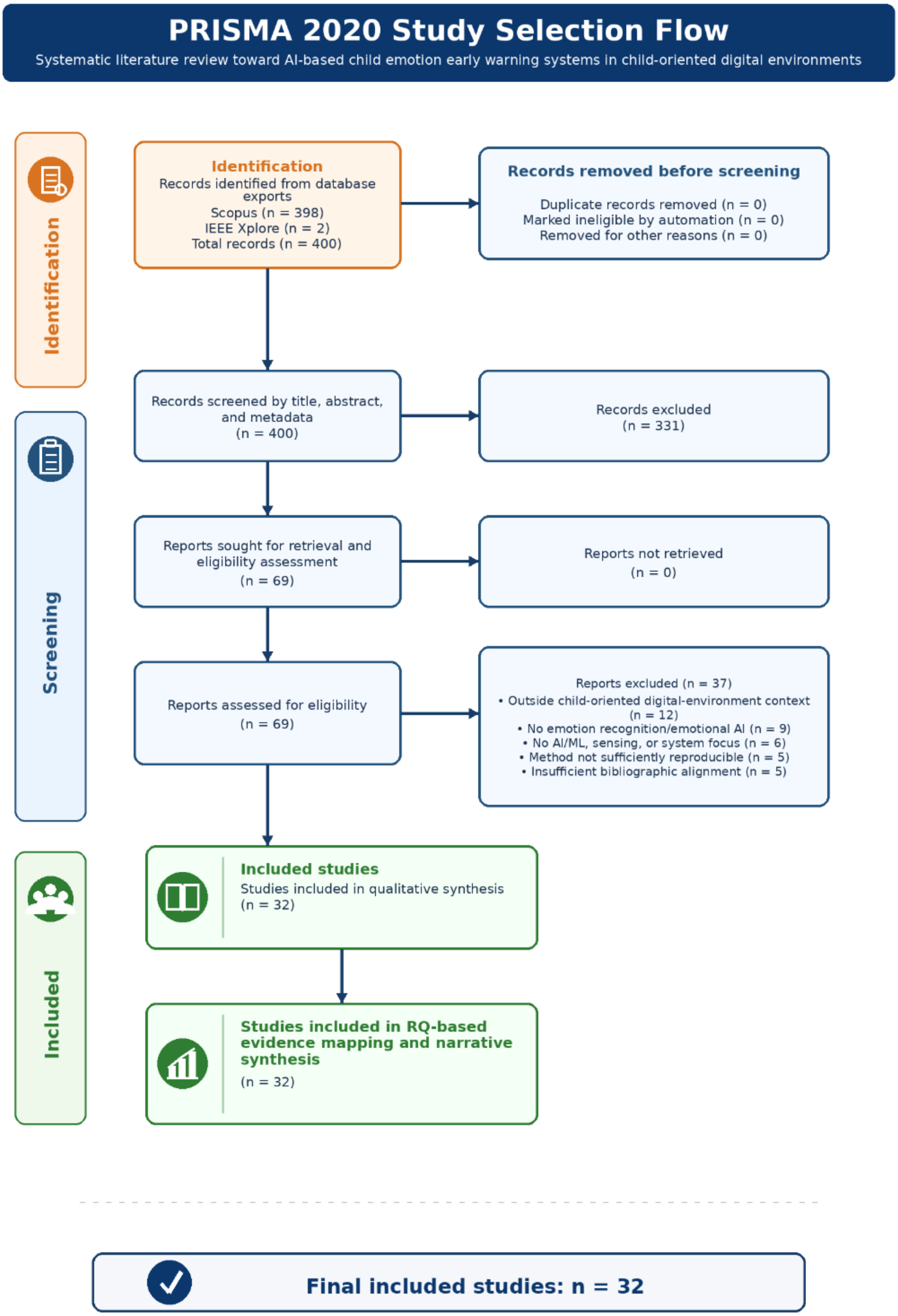

Figure 1. Core PRISMA 2020 study-selection flow

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 5. Evidence maps by research question

RQ	Analytical focus	Core evidence/counts	Key finding
RQ1	Sensing modalities	Facial/visual n=19; physiological/sensors n=7; speech/audio n=7; multimodal n=6; gameplay/interaction n=5; text/dialogue n=2; gesture/pose n=2; gaze n=1	Visual sensing dominates; multimodal evidence remains comparatively limited.
RQ2	AI methods/model families	CNN/deep CNN n=12; multimodal/fusion n=6; Transformer/ViT n=5; acceptance/governance n=5; classical ML n=4; XAI n=3; real-time/IoT/fog n=2; generative/AR n=1	CNN-based methods dominate, while Transformer, XAI, and real-time implementations remain emerging.
RQ3	Population and dataset context	Children/young children n=21; general/all-age n=10; ASD/ADHD n=10; youth/students n=6; parents/caregivers n=5	Evidence is heterogeneous, with substantial neurodevelopmental and mixed-age representation and limited generalizability.
RQ4	Alert translation and system design	Emotion-responsive systems and caregiver-facing alert translation: n = 12	Evidence for emotion recognition is stronger than evidence for operational caregiver-facing alert translation.
RQ5	Socio-technical factors	Privacy/security n=9; ethics/governance n=7; acceptance/intention n=5; parental/caregiver role n=5; transparency/XAI n=4; trust n=2;	Privacy and governance receive more attention than empirically validated caregiver acceptance or trust.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

RQ	Analytical focus	Core evidence/counts	Key finding
		accuracy/reliability n=2; risk/moral risk n=2	

Note. Counts are non-exclusive; a single study may contribute to more than one category within or across RQs. All counts derive exclusively from the 32 core included studies.

Child-focused work on gaming [20], engagement recognition [29], and pediatric mobile classification [27] illustrates the prominence of face-based analysis. However, the evidence does not support treating facial cues as sufficient for every child-oriented digital setting. A child's face may be partially occluded, weakly expressive, or difficult to interpret without information about voice, posture, interaction difficulty, or child-oriented interaction context. Multimodal designs may improve interpretation in such cases, but they also introduce additional requirements for synchronisation, consent, secure data handling, and computational efficiency.

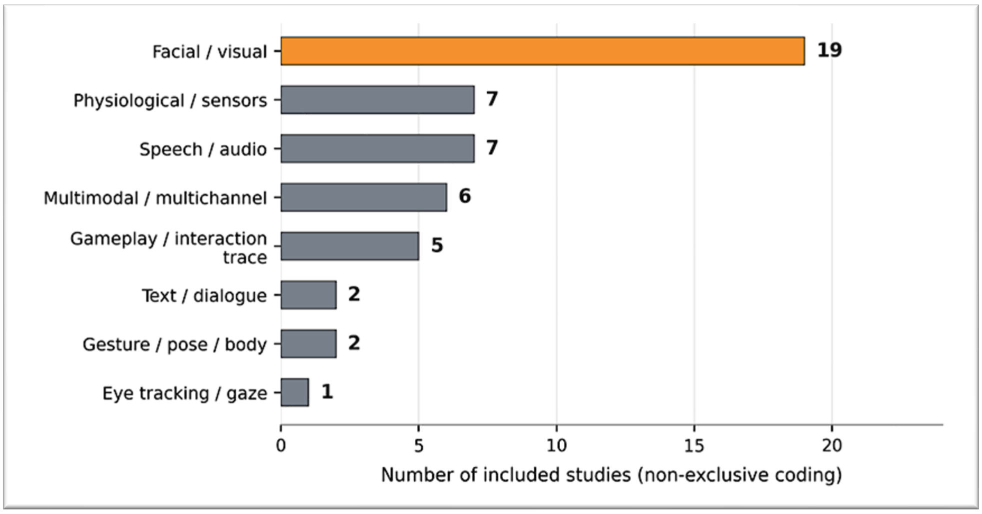

Figure 2. Evidence map of sensing modalities in the core PRISMA corpus (n = 32; non-exclusive counts).

3.2. RQ2. AI methods and model families

Figure 3 shows that CNN or deep CNN models were the most common method family, appearing in 12 studies. Multimodal or fusion models appeared in six studies, Transformer or Vision Transformer approaches in five studies, and acceptance or governance frameworks in five studies. Classical machine-learning methods, including SVM and

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

decision-tree approaches, appeared in four studies, while explainability-focused studies appeared in three studies. Real-time, IoT, or fog-based implementations were found in two studies, and one study addressed a generative or AR-assisted system. These categories are non-exclusive and should not be interpreted as a standardised benchmark comparison across identical datasets.

The dominance of CNN-based approaches reflects their maturity for image-based facial-expression recognition and transfer learning. Transformer-based methods are emerging, but the reviewed studies use different datasets, sample sizes, emotion labels, and evaluation procedures; therefore, the corpus does not justify a general claim that one architecture consistently outperforms another. Santoso et al. [20], Radocaj and Martinovic [19], and Banerjee et al. [27] demonstrate the relevance of child-specific modelling, while the broader evidence suggests that model selection should consider dataset suitability, class balance, inference latency, and interpretability alongside predictive performance.

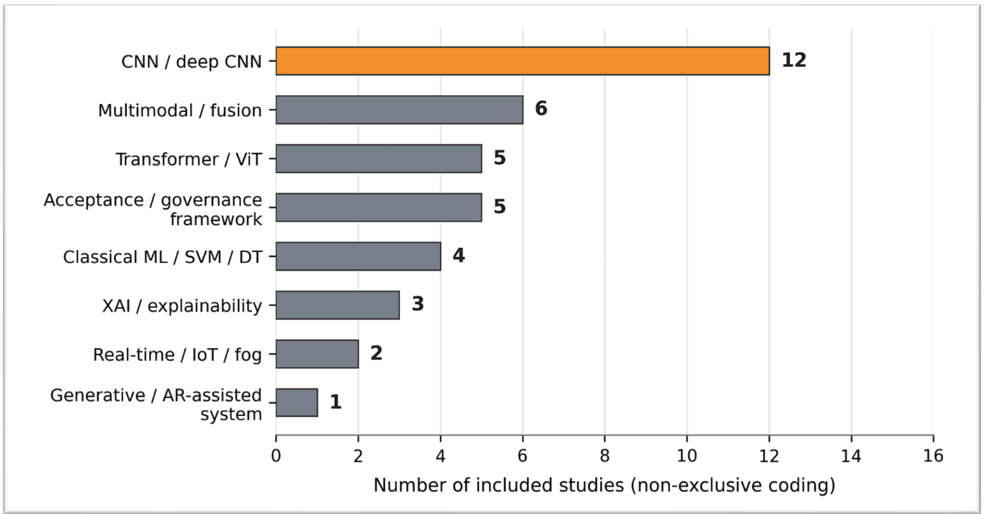

Figure 3. Evidence map of AI method and system families in the core PRISMA corpus (n = 32; non-exclusive counts).

3.3. RQ3. Dataset and population focus

Figure 4 shows that the core corpus is child-oriented but methodologically heterogeneous. Of the 32 included studies, 21 were coded as involving children or young children; however, 10 studies also included general or all-age populations, 10 focused on ASD or ADHD contexts, six involved youth or students, and five involved parents or

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

caregivers. Because these categories are non-exclusive, the figure does not indicate a single representative child dataset. Instead, it shows that the available evidence is drawn from clinical, assistive, educational, and general affective-computing settings, each with different interaction conditions. For example, Santoso et al. [20] examined children's facial expressions in a gaming context, whereas Radocaj and Martinovic [19] focused on autistic children. These studies are relevant to child emotion recognition, but their findings cannot be assumed to represent spontaneous emotional expression across all child-oriented digital settings.

Dataset suitability, therefore, remains a central limitation. Banerjee et al. [27] demonstrated the value of a pediatric facial-expression classifier for mobile devices, while Kim et al. [28] showed that facial emotion recognition performance may differ across age groups. For an early warning system, model evaluation should report the child age range, data-collection setting, emotion-class distribution, annotation procedure, and performance across relevant subgroups. The proposed system should consequently treat predicted emotion labels as probabilistic signals for caregiver awareness, not as definitive statements about a child's psychological condition.

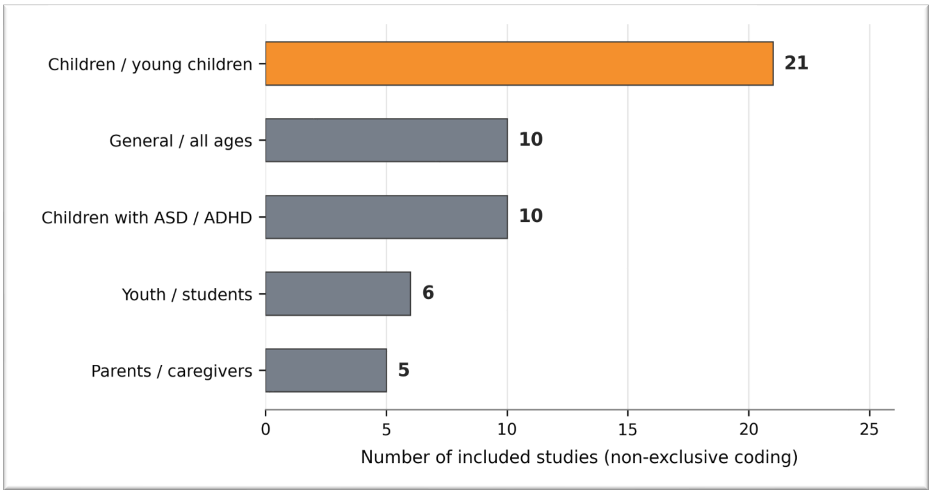

Figure 4. Population and dataset contexts in the core PRISMA corpus (n = 32; non-exclusive coding).

3.4. RQ4. Early warning translation and emotion-to-alert mapping

Figure 5 should be interpreted as a proposed decision-support structure rather than as a frequency-based result. The reviewed studies provide evidence that emotion

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

recognition can be incorporated into real-time feedback, intelligent tutoring, participatory serious games, and child-oriented interactive systems [22], [26], [30]. However, the core corpus does not establish a common threshold for when a detected emotion should lead to logging, monitoring, caregiver notification, or intervention. Model accuracy alone is insufficient for this purpose because the significance of an emotion may depend on its duration, recurrence, interaction context, and prediction confidence. Ethical guidance for affect detection and AI in games also supports a cautious approach in which system outputs are explainable and do not overstate their certainty [31], [32].

Based on this gap, the present study proposes a three-level alert mechanism. Happy and neutral are classified as No Alert states because they do not ordinarily require caregiver notification. Sad and fear are placed in the Watchlist category because their meaning may depend on duration, recurrence, model confidence, and the child-oriented digital context. Fear may warrant escalation when it persists, recurs, or appears with additional signs of avoidance or distress, but a single fear label in a game, AR, or learning context may also reflect surprise, challenge, or play-related tension. Angry and crying are classified as High Alert candidates because they may indicate more immediate distress and may justify caregiver notification when the signal persists, recurs, or is detected with high confidence. In practical implementation, alert sensitivity should also be configurable according to caregiver preferences, provided that such configuration does not override essential child-safety and privacy safeguards. Crying should also be interpreted cautiously because its meaning depends on age, developmental stage, duration, context, and caregiver knowledge of the child. This mapping is a conceptual design outcome of the review, not a universal or clinical standard; its notification rules and caregiver response options require empirical validation.

Figure 5. Proposed design mapping of AI-inferred emotions to caregiver alert categories.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Table 6. Proposed operational emotion-to-alert mapping.

AI inference label	Alert category	Decision meaning	System action
Happy	No Alert	Positive or comfortable interaction signal	No notification; optional logging
Neutral	No Alert	Baseline, non-distress signal	No notification; continue monitoring
Sad	Watchlist	Possible low-intensity distress	Log and monitor persistence, recurrence, and confidence
Fear	Watchlist	Possible situational distress	Log and assess repetition, duration, and interaction context
Angry	High Alert	Possible active distress	Notify the caregiver when persistent, repeated, or high-confidence
Crying	High Alert	Observable high-distress signal	Send priority notification for caregiver review

Note. This mapping supports caregiver awareness and is not a clinical diagnosis. Alert escalation should consider duration, recurrence, prediction confidence, multimodal agreement, interaction context, child age or developmental stage, and caregiver preferences or configurable alert settings, rather than relying on a single predicted emotion label.

3.5. RQ5. Trust, privacy, ethics, and caregiver or parental acceptance

Figure 6 shows that privacy and security were the most frequently coded socio-technical factors, appearing in nine studies. Ethics and governance appeared in seven studies, while acceptance or intention and parental or caregiver roles were each represented in five studies. Transparency or explainability appeared in four studies, whereas trustworthiness, accuracy or reliability, and moral risk were each identified in two studies. These non-exclusive counts do not indicate causal strength; rather, they show that child emotion-recognition systems are discussed not only as technical tools but also as systems

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

involving governance, caregiver involvement, and responsible data practices. McStay and Rosner [12] highlight ethical and governance concerns in emotional AI for children's devices, while Ho et al. [13] relate moral-risk awareness to acceptance of emotional AI. Afroogh et al. [17] further describe trust in AI as dependent on both technical reliability and institutional conditions.

For the proposed early warning system, acceptance should not be assumed to follow from prediction accuracy alone. The system should provide understandable notifications, parental or caregiver control over data collection and alert settings, clear consent procedures, child assent where appropriate, data minimization, retention limits, secure storage, transparent deletion procedures, and a visible explanation of uncertainty when emotion predictions are displayed. These elements can be examined in subsequent empirical research through perceived usefulness, perceived accuracy, trust, privacy concern, ethical risk, satisfaction, and adoption intention. Such a design positions the system as a privacy-sensitive service that supports caregiver awareness rather than as a surveillance tool or automated diagnostic system.

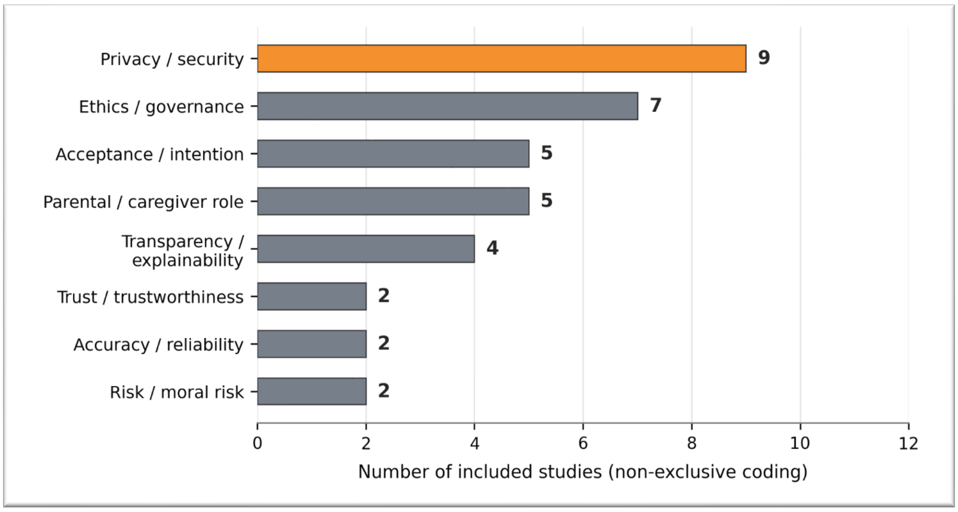

Figure 6. Trust, privacy, ethics, and caregiver-related factors in the core PRISMA corpus (n = 32; non-exclusive coding).

Supplemental interpretive evidence is reported in Appendix A and is not included in the PRISMA count, RQ-specific frequencies, or core quality-appraisal summary.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

3.6. Socio-Technical Risks in Caregiver Alert Translation

Translating AI-inferred emotions into caregiver-facing alerts introduces socio-technical risks that go beyond classification performance. False positives may create unnecessary caregiver anxiety, over-monitoring, or alert fatigue, especially when alerts are triggered by short-lived emotional expressions. False negatives may be more serious because persistent distress may be missed when the model fails to detect subtle, occluded, or context-dependent emotional cues. Therefore, alert logic should not depend on a single predicted label but should consider duration, recurrence, confidence, multimodal agreement, and interaction context.

Interpretation also depends on cultural norms, developmental stage, neurodiversity, and the specific digital activity. A facial expression labelled as fear, sadness, or anger may have different meanings in a game, therapeutic task, online-learning activity, or social interaction. For autistic children and other neurodiverse populations, emotional expression may differ from typical benchmark assumptions; therefore, system outputs should be treated as probabilistic caregiver-awareness signals rather than definitive judgments about a child's emotional state.

Table 7. Potential risks of caregiver-facing emotion alerts

Risk/Error	Example	Potential harm	Mitigation
False positive	Short-lived fear during a game is interpreted as distress	unnecessary interruption, caregiver anxiety, alert fatigue, excessive monitoring	persistence threshold, confidence filtering, multimodal confirmation, caregiver review
False negative	Persistent distress is not detected	delayed caregiver support	recurrence analysis, multiple modalities, periodic summaries
Context misclassification	Game challenge interpreted as psychological distress	inappropriate escalation	interaction-context awareness and caregiver interpretation

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

Risk/Error	Example	Potential harm	Mitigation
Developmental misinterpretation	Crying treated identically across ages	overgeneralized alert response	age/development-sensitive interpretation

Implementation risks should also be considered. Device placement, camera angle, lighting, background noise, microphone quality, and interaction distance can affect model reliability. Device availability, computational capability, network connectivity, and variation in household hardware may also affect inference latency, accessibility, and the feasibility of edge-based processing. Child-oriented systems should adopt privacy-by-design principles, including data minimisation, retention limits, local or edge-based processing where feasible, secure storage, transparent deletion procedures, parental consent, and child assent where appropriate. Caregiver-facing alerts should be configurable and proportionate to prevent alert overload, unnecessary anxiety, and over-monitoring.

Beyond parental consent, child-oriented emotional AI should respect children's rights, dignity, age-appropriate autonomy, and evolving capacity to participate in decisions about digital monitoring. Where developmentally appropriate, child assent should complement parental consent, and children should have meaningful opportunities to understand, question, pause, or disengage from continuous emotion monitoring. These safeguards are essential to ensure that caregiver-facing alerts support child well-being without normalizing excessive surveillance.

Based on the core evidence map, the proposed framework is positioned as a socio-technical direction toward AI-based child emotion early warning systems in child-oriented digital environments. The system begins with child-oriented digital interactions, captures emotional signals, performs AI-based inference, translates inferred states into caregiver-facing alert categories, and communicates the result through explainable and privacy-sensitive mechanisms. At the alert-decision layer, happy and neutral are mapped to No Alert, sad and fear to Watchlist, and angry and crying to High Alert. This mapping is intended as a risk-communication mechanism that supports caregiver awareness; it is not a clinical diagnosis.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

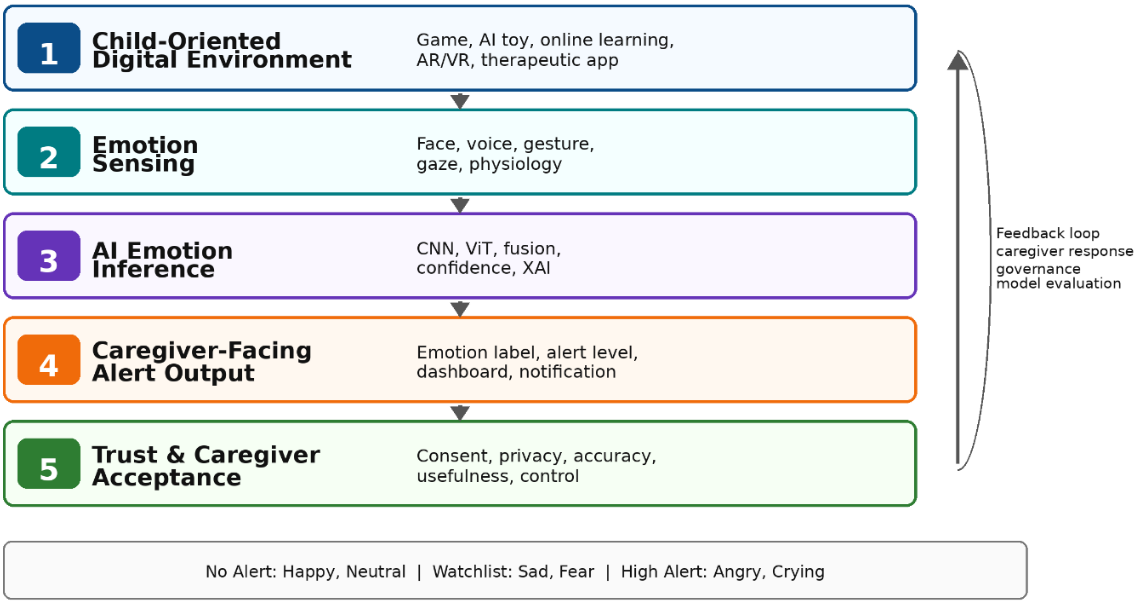

Figure 7. Proposed socio-technical framework toward child emotion early warning systems in child-oriented digital environments, synthesised from the core evidence map.

The proposed framework has several implications. First, technical model performance should be evaluated alongside ecological validity, interpretability, uncertainty communication, and usability. Second, child-centred systems require privacy by design, data minimisation, secure storage, parental or caregiver control, and transparent consent procedures. Third, the caregiver-acceptance layer can be operationalised in subsequent empirical research through constructs such as perceived usefulness, perceived ease of use, trust, perceived accuracy, privacy concern, ethical risk, satisfaction, and adoption intention. Finally, future systems should avoid treating children as passive data sources and should instead prioritise developmental appropriateness, caregiver mediation, and responsible governance.

4. CONCLUSION

This systematic literature review shows that work toward AI-based child emotion early warning systems in child-oriented digital environments requires more than accurate emotion classification. The core evidence from 32 studies indicates that facial/visual sensing and CNN-based methods are the most common technical approaches, while child-

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

specific datasets, ecological validity, multimodal integration, and age-aware evaluation remain limited. The strongest gap is not emotion recognition itself, but the translation of uncertain emotion predictions into understandable, proportionate, and caregiver-facing alerts. The review proposes a socio-technical framework that links digital interaction context, emotion sensing, AI inference, alert translation, and trust or caregiver acceptance. The proposed happy/neutral–No Alert, sad/fear–Watchlist, and angry/crying–High Alert mapping should be treated as a design proposition, not as a validated clinical or operational rule.

The study is limited by the scope of the core databases, the heterogeneity of included contexts, and the incomplete status of the supplemental audit, which is reported separately in Appendix A. The mandatory socio-technical search block may also have reduced retrieval of purely technical child emotion-recognition studies that did not explicitly report trust, privacy, ethics, acceptance, or related concepts in searchable metadata. In addition, screening and initial extraction were conducted by one reviewer and subsequently validated through co-author consensus; therefore, formal inter-rater agreement was not estimated. The framework is not ready for deployment without parent/caregiver validation, child-safety review, privacy assessment, and pilot testing. Future research should develop and evaluate prototypes with parents, caregivers, teachers, therapists, and child-development experts. Empirical work should examine usability, alert thresholds, false positives and false negatives, caregiver fatigue, child assent, privacy safeguards, and the effect of age, culture, neurodiversity, and interaction context on alert interpretation.

DATA AVAILABILITY STATEMENT

Supplementary Table S1 contains the study-level characteristics and extraction matrix for the 32 core included studies. Supplementary Table S2 provides the complete C1–C7 study-level quality-appraisal matrix. Supplementary File S3 contains the structured data-extraction form, Supplementary File S4 contains the seven-criterion quality-appraisal codebook, and Supplementary Table S5 reports full-text exclusion decisions and reasons. Core RIS exports, database-provenance logs, screening trackers, DOI-audit records, and other bibliographic audit materials are retained by the authors and may be made available upon reasonable request, subject to database licensing and copyright restrictions.

ISI Journal of Information Systems and Informatics

Published By Asosiasi Doktor Sistem Informasi Indonesia

ACKNOWLEDGMENT

The authors acknowledge the academic discussions and scholarly support within the Doctoral Program of Information Systems, Universitas Diponegoro. These exchanges helped refine the scope of this review and strengthen its socio-technical perspective on AI-based child emotion early warning systems in child-oriented digital environments.

REFERENCES

[1] S. K. Khare, V. Blanes-Vidal, E. S. Nadimi, and U. R. Acharya, "Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations," *Information Fusion*, vol. 102, 2024, doi: 10.1016/j.inffus.2023.102019.

[2] A. O. R. Vistorte, A. Deroncele-Acosta, J. L. M. Ayala, A. Barrasa, C. López-Granero, and M. Martí-González, "Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review," *Front. Psychol.*, vol. 15, Jun. 2024, doi: 10.3389/fpsyg.2024.1387089.

[3] S. Yu, A. Androsov, H. Yan, and Y. Chen, "Bridging computer and education sciences: A systematic review of automated emotion recognition in online learning environments," *Comput. Educ.*, vol. 220, 2024, doi: 10.1016/j.compedu.2024.105111.

[4] S. Kalateh, L. A. Estrada-Jiménez, S. Nikghadam-Hojjati, and J. Barata, "A systematic review on multimodal emotion recognition: Building blocks, current state, applications, and challenges," *IEEE Access*, vol. 12, pp. 103976–104019, 2024, doi: 10.1109/ACCESS.2024.3430850.

[5] R. Pereira *et al.*, "Systematic review of emotion detection with computer vision and deep learning," *Sensors*, vol. 24, no. 11, 2024, doi: 10.3390/s24113484.

[6] A. Landowska *et al.*, "Automatic emotion recognition in children with autism: A systematic literature review," *Sensors*, vol. 22, no. 4, 2022, doi: 10.3390/s22041649.

[7] M. A. Rashidan *et al.*, "Technology-assisted emotion recognition for autism spectrum disorder (ASD) children: A systematic literature review," *IEEE Access*, vol. 9, pp. 33638–33653, 2021, doi: 10.1109/ACCESS.2021.3060753.

[8] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n71.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[9] M. J. Page *et al.*, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *BMJ*, vol. 372, 2021, doi: 10.1136/bmj.n160.

[10] M. Fan, S. Jin, J. Fan, W. Guo, and X. Chen, "EmoLand: Utilizing narrative animations, multilevel games, and affective computing to foster emotional development in children with autism spectrum disorder," *Int. J. Hum. Comput. Stud.*, vol. 199, 2025, doi: 10.1016/j.ijhcs.2025.103486.

[11] D. Sosnowski *et al.*, "Brief report: A novel digital therapeutic that combines applied behavior analysis with gaze-contingent eye tracking to improve emotion recognition in children with autism spectrum disorder," *J. Autism Dev. Disord.*, vol. 52, pp. 2357–2366, 2022, doi: 10.1007/s10803-021-05101-w.

[12] A. McStay and G. L. Rosner, "Emotional artificial intelligence in children's toys and devices: Ethics, governance and practical remedies," *Big Data Soc.*, vol. 8, 2021, doi: 10.1177/2053951721994877.

[13] M.-T. Ho, P. Mantello, and Q.-H. Vuong, "Emotional AI in education and toys: Investigating moral risk awareness in the acceptance of AI technologies from a cross-sectional survey of the Japanese population," *Heliyon*, vol. 10, no. 16, 2024, doi: 10.1016/j.heliyon.2024.e36251.

[14] A. Karran *et al.*, "Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education," *NPJ Sci. Learn.*, vol. 10, 2025, doi: 10.1038/s41539-025-00333-2.

[15] I. R. Berson, M. J. Berson, and W. Luo, "Innovating responsibly: Ethical considerations for AI in early childhood education," *AI, Brain and Child*, vol. 1, 2025, doi: 10.1007/s44436-025-00003-5.

[16] W. Xiao and A. Goncalves, "Intelligent toys, complex questions: A literature review of artificial intelligence in children's toys and devices," *Big Data Soc.*, vol. 12, 2025, doi: 10.1177/20539517251389860.

[17] S. Afroogh, A. Akbari, E. Malone, M. Kargar, and H. Alambeigi, "Trust in AI: Progress, challenges, and future directions," *Humanit. Soc. Sci. Commun.*, vol. 11, 2024, doi: 10.1057/s41599-024-04044-8.

[18] A. Shrestha *et al.*, *Navigating AI to Unpack Youth Privacy Concerns: An In-Depth Exploration and Systematic Review*. 2024. doi: 10.48550/arXiv.2412.16369.

[19] P. Radocaj and G. Martinovic, "Emotion recognition in autistic children through facial expressions using advanced deep learning architectures," *Applied Sciences*, vol. 15, no. 17, 2025, doi: 10.3390/app15179555.

[20] H. Santoso, G. F. Soares, and C. M. Angelo, "Convolutional neural network-based recognition of children's facial expressions in response to gaming," *International Journal of Advanced Science Computing and Engineering*, vol. 6, no. 3, pp. 132–137, 2024.

[21] P. Soleiman *et al.*, "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," *Front. Robot. AI*, vol. 10, 2023, doi: 10.3389/frobt.2023.1088582.

[22] X. Ruan, C. Palansuriya, A. Constantin, and K. Tsiakas, "Supporting Children's Metacognition with a Facial Emotion Recognition based Intelligent Tutor System," in *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 2023, pp. 502–506. doi: 10.1145/3585088.3593882.

[23] A. R. Junaidi, M. Irvan, M. N. Jauhari, A. H. Putri, and M. F. Restuti, "Emotion-Selfie: A Tool for Enhancing Emotional Expression Skills in Children with Autism," in *Proceedings - 2025 11th International Conference on Education and Technology: Personalized AI for Digital Well-being: Bridging Technology and Humanity, ICET 2025*, IEEE, 2025, pp. 220–225. doi: 10.1109/ICET67257.2025.11290889.

[24] A. R. Junaidi, M. Irvan, J. Yuwono, D. A. Dewantoro, Y. Pramesti, and E. P. Puspitasari, "Usability testing of emotion recognition assistance application for children with autism spectrum disorder," in *Proceedings - International Conference on Education and Technology, ICET*, IEEE, 2022, pp. 309–314. doi: 10.1109/ICET56879.2022.9990633.

[25] R. Pavez, J. Diaz, J. Arango-Lopez, D. Ahumada, C. Mendez-Sandoval, and F. Moreira, "Emo-mirror: A proposal to support emotion recognition in children with autism spectrum disorders," *Neural Comput. Appl.*, vol. 35, no. 11, pp. 7913–7924, 2023, doi: 10.1007/s00521-021-06592-5.

[26] X. Ruan, C. Palansuriya, and A. Constantin, "Real-time feedback based on emotion recognition for improving children's metacognitive monitoring skill," in *Proceedings of Interaction Design and Children*, 2022, pp. 672–675.

[27] A. Banerjee, O. C. Mutlu, A. Kline, S. Surabhi, P. Washington, and D. P. Wall, "Training and profiling a pediatric facial expression classifier for children on mobile devices: Machine learning study," *JMIR Form. Res.*, vol. 7, 2023, doi: 10.2196/39917.

[28] E. Kim, D. Bryant, D. Srikanth, and A. Howard, "Age bias in emotion detection: An analysis of facial emotion recognition performance on young, middle-aged, and older adults," in *AIES 2021 - Proceedings of the 2021 AAAI/ACM Conference on AI,*

Ethics, and Society, Association for Computing Machinery, 2021, pp. 638–644. doi: 10.1145/3461702.3462609.

[29] W.-H. Yun, D. Lee, C. Park, J. Kim, and J. Kim, "Automatic recognition of children engagement from facial video using convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 4, pp. 696–707, 2020, doi: 10.1109/TAFFC.2018.2834350.

[30] H. K. H. Abd El-Sattar, M. Omar, and H. Mohamady, "Developing a participatory research framework through serious games to promote learning for children with autism," *Front. Educ. (Lausanne)*, vol. 9, 2024, doi: 10.3389/feduc.2024.1453327.

[31] A. M. Banzon, J. Beever, and M. Taub, "Facial expression recognition in classrooms: Ethical considerations and proposed guidelines for affect detection in educational settings," *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 93–104, 2024, doi: 10.1109/TAFFC.2023.3275624.

[32] D. Melhart, J. Togelius, B. Mikkelsen, C. Holmgard, and G. N. Yannakakis, "The ethics of AI in games," *IEEE Trans. Affect. Comput.*, vol. 15, no. 1, pp. 79–92, 2024, doi: 10.1109/TAFFC.2023.3276425.

[33] S. Cano, C. S. González, R. M. Gil-Iranzo, and S. Albiol-Pérez, "Affective communication for socially assistive robots (SARs) for children with autism spectrum disorder: A systematic review," *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21155166.

[34] S. Kirst *et al.*, "Fostering socio-emotional competencies in children on the autism spectrum using a parent-assisted serious game: A multicenter randomized controlled trial," *Behaviour Research and Therapy*, vol. 152, 2022, doi: 10.1016/j.brat.2022.104068.

[35] H. Drimalla, I. Baskow, B. Behnia, S. Roepke, and I. Dziobek, "Imitation and recognition of facial emotions in autism: A computer vision approach," *Mol. Autism*, vol. 12, no. 1, 2021, doi: 10.1186/s13229-021-00430-0.

[36] O. Baños *et al.*, "Sensing technologies and machine learning methods for emotion recognition in autism: Systematic review," *Int. J. Med. Inform.*, vol. 187, 2024, doi: 10.1016/j.ijmedinf.2024.105469.

[37] D. R. Faria, A. L. Godkin, and P. P. da Silva Ayrosa, "Advancing emotionally aware child-robot interaction with biophysical data and insight-driven affective computing," *Sensors*, vol. 25, no. 4, 2025, doi: 10.3390/s25041161.

[38] F. M. Talaat, "Real-time facial emotion recognition system among children with autism based on deep learning and IoT," *Neural Comput. Appl.*, vol. 35, no. 17, pp. 12717–12728, 2023, doi: 10.1007/s00521-023-08372-9.

[39] A. Sousa, K. Young, M. d'Aquin, M. Zarrouk, and J. Holloway, "Introducing CALMED: Multimodal annotated dataset for emotion detection in children with autism," in *Universal Access in Human-Computer Interaction*, vol. 14020, in Lecture Notes in Computer Science, vol. 14020. , Cham, Switzerland: Springer, 2023, pp. 657–677. doi: 10.1007/978-3-031-35681-0_43.

[40] S. K. B. Sangeetha, R. S. K. Boddu, A. Bhaumik, S. K. Mathivanan, and U. Moorthy, "Spatiotemporal multimodal emotion recognition using temporal video sequences and pose features for child emotion classification," *Sci. Rep.*, vol. 15, 2025, doi: 10.1038/s41598-025-25813-8.

[41] Z. B. Turabee *et al.*, "A privacy-preserving approach for engagement and meltdown detection in children with autism using machine learning models," in *Neural Information Processing: 31st International Conference, ICONIP 2024, Proceedings, Part IV*, vol. 2285, in Communications in Computer and Information Science, vol. 2285. , Singapore: Springer, 2025, pp. 274–286. doi: 10.1007/978-981-96-6957-8_19.

[42] B. Zupan and M. Eskritt, "Facial and vocal emotion recognition in adolescence: A systematic review," *Adolesc. Res. Rev.*, vol. 9, no. 2, pp. 253–277, 2024, doi: 10.1007/s40894-023-00219-7.

[43] S. R. Gudivaka *et al.*, "Speech emotion recognition in adults and children: A comprehensive review of traditional features and raw waveform models," *Int. J. Speech Technol.*, vol. 29, no. 1, 2026, doi: 10.1007/s10772-025-10229-6.

[44] T. A. D'Amelio *et al.*, "Emotion recognition systems with electrodermal activity: From affective science to affective computing," *Neurocomputing*, vol. 651, 2025, doi: 10.1016/j.neucom.2025.130831.

[45] H. F. T. Al-Saadawi, B. Das, and R. Das, "A systematic review of trimodal affective computing approaches: Text, audio, and visual integration in emotion recognition and sentiment analysis," *Expert Syst. Appl.*, vol. 255, 2024, doi: 10.1016/j.eswa.2024.124852.

[46] S. C. Leong, Y. M. Tang, C. H. Lai, and C. K. M. Lee, "Facial expression and body gesture emotion recognition: A systematic review on the use of visual data in affective computing," *Comput. Sci. Rev.*, vol. 48, 2023, doi: 10.1016/j.cosrev.2023.100545.

[47] S. Kadyr and T. Chinibayeva, "Affective computing methods for simulation of action scenarios in video games," *Procedia Comput. Sci.*, vol. 231, pp. 341–346, 2024, doi: 10.1016/j.procs.2023.12.214.

[48] R. E. Lee *et al.*, "Affective computing in youth psychopathology: Protocol for the Adolescent Communication of Emotions Study," *Pers. Med. Psychiatry*, vol. 55, 2026, doi: 10.1016/j.pmip.2026.100181.

[49] E. L.-C. Law, S. Soleimani, D. Watkins, and J. Barwick, "Automatic voice emotion recognition of child-parent conversations in natural settings," *Behaviour & Information Technology*, vol. 40, no. 11, pp. 1072–1089, 2021, doi: 10.1080/0144929X.2020.1741684.

[50] Y. Lyu *et al.*, "EMooly: Supporting autistic children in collaborative social-emotional learning with caregiver participation through interactive AI-infused and AR activities," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, pp. 1–36, 2024, doi: 10.1145/3699738.

[51] L. Qi, J.-W. Ni, G. Dong, T. Sun, and J.-W. Zhang, "Cortical hemodynamic responses and deep learning models of emotional face processing in preschool children with autism spectrum disorder: a fNIRS study," *Front. Psychiatry*, vol. 16, Art. no. 1703302, 2026, doi: 10.3389/fpsyt.2025.1703302.

[52] N. G. Wall, O. Smith, L. Campbell, C. Loughland, and U. Schall, "Using EEG and eye tracking to evaluate an emotion recognition iPad app for autistic children," *Clin. EEG Neurosci.*, vol. 57, no. 2, pp. 130–140, 2026, doi: 10.1177/15500594251362402.

[53] R. Dadi, S. Gaddam, and A. Vamshikirishna, "Children Emotion Detection Deep Learning: A Comparative Study," in *Proc. 7th Int. Conf. Innovative Data Communication Technologies and Application (ICIDCA)*, 2025, doi: 10.1109/ICIDCA66325.2025.11280509.

[54] D. Madhusoodanan, N. Sagar, V. R., H. Santhosh, and S. Palaniswamy, "Facial Emotion Detection in Autistic Children: A Deep Learning Approach," in *Proc. 2025 Int. Conf. Electronics, Computing, Communication and Control Technology (ICECCC)*, 2025, doi: 10.1109/ICECCC65144.2025.11063965.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

APPENDIX A. Reproducibility, Source-Provenance, And Supplemental Screening Audit

The core PRISMA evidence map was based on 400 records from Scopus (n = 398) and IEEE Xplore (n = 2), retrieved on 17 June 2026. After screening, full-text assessment, and quality appraisal, 32 studies were retained for the core synthesis and RQ-specific evidence maps. A separate audit identified 2,037 records from SpringerLink, PubMed, ScienceDirect, and Emerald. Source provenance was verified through record-level URLs and RIS identifiers, while duplicates were matched using normalised DOI strings or, where unavailable, titles and publication years. The combined pool contained 2,437 records; after removing 105 exact duplicates, 2,332 unique records remained. Initial screening produced 241 provisional records, consolidated into 232 DOI-level candidates. Batch 1 reviewed 13 priority candidates using the same eligibility and quality criteria as the core corpus. Six studies were provisionally eligible, six were excluded, and one overlapped with the core corpus. Because 219 candidates remain pending equivalent review, the supplemental audit does not change the PRISMA count or RQ-specific figures; it is retained to document search transparency and support future review expansion.

Supplementary database enrichment and interpretive validation

The supplemental audit was conducted to improve source transparency and identify potentially relevant studies beyond the core Scopus and IEEE Xplore corpus. It was not intended to redefine the PRISMA evidence base. Therefore, all quantitative findings reported in Figures 1–6 remain derived from the 32 studies included in the core corpus. Source-level title-and-metadata screening identified 241 provisional records across SpringerLink, PubMed, ScienceDirect, and Emerald. DOI consolidation reduced these records to 232 unique DOI-level candidates. A first priority batch of 13 candidates completed a full-text eligibility assessment and a seven-criterion quality appraisal. Six studies were provisionally eligible, six were excluded, and one duplicated a study already included in the core corpus. The provisional supplemental studies are retained only as contextual interpretive support because 219 DOI-level candidates remain pending review. Their findings are therefore used to support the interpretation of multimodal sensing, child-specific datasets, real-time systems, and privacy-sensitive design, but they are not included in the PRISMA flow, evidence-map frequencies, or RQ-specific quantitative claims.

Table 8. Supplemental source-provenance and staged screening status.

Source	Export date and raw records	Screening outcome	Full-text appraisal status
SpringerLink	3 May 2026 Raw: 1,000	Two internal duplicates removed; 998 unique records screened; 141 provisional candidates identified.	Three records entered Batch 1. Two were provisionally eligible, and one overlapped with the core corpus. None were added to the core PRISMA corpus.
PubMed	11 April 2026 Raw: 977	977 unique records screened; 86 provisional candidates identified.	Eleven records entered Batch 1. Five were provisionally eligible, and six were excluded. One provisional included was co-indexed with SpringerLink. None were added to the core PRISMA corpus.
ScienceDirect	2 June 2026 Raw: 25	25 unique records screened; nine provisional candidates identified.	No candidate from this source entered Batch 1. Records remain pending full-text eligibility screening and quality appraisal.
Emerald	2 June 2026 Raw: 35	35 unique records screened; five provisional candidates identified.	No candidate from this source entered Batch 1. Records remain pending full-text eligibility screening and quality appraisal.

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

Source	Export date and raw records	Screening outcome	Full-text appraisal status
Combined audit boundary	Core plus supplemental raw records: 2,437	After removal of 105 exact bibliographic duplicates, 2,332 unique records remained. DOI consolidation yielded 232 unique supplemental candidates.	13 unique DOI-level candidates: six provisional supplemental includes, six exclusions, and one core overlap. The remaining 219 candidates are pending equivalent review. The core corpus remains n = 32.

Supplemental full-text appraisal: Batch 1

The first supplemental full-text batch provides a protocol-constrained robustness check rather than a new quantitative synthesis. Six studies met the eligibility criteria and received high-quality scores (mean 12.7/14). They extend contextual interpretive support for child–robot interaction, pediatric mobile facial-expression classification, multimodal video–pose modelling, real-time IoT systems, fNIRS-based deep learning, and game-like iPad interventions. Four of these sources are already represented in the core reference base [27], [37], [38], [40], whereas the two additional full-text-appraised sources are provided as [51] and [52]. Because 219 remaining candidates have not completed harmonized full-text review and structured data extraction, these findings do not alter the core evidence-map frequencies.

Table 9. Provisional supplemental evidence from the completed full-text appraisal batch

DOI	Study Focus	Relevant RQ(s)	QA score	Status
10.3390/s25041161	Child–robot interaction; facial and speech emotion signals.	RQ1, RQ2, RQ4, RQ5	**12/14**	**Provisional include**

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

10.2196/39917	Pediatric facial-expression classifier for mobile devices.	RQ1, RQ2, RQ3, RQ5	**14/14**	**Provisional include**
10.1038/s41598-025-25813-8	Child multimodal video–pose emotion classification.	RQ1, RQ2, RQ3, RQ4	**13/14**	**Provisional include**
10.1007/s00521-023-08372-9	Real-time facial emotion recognition using deep learning and IoT.	RQ1, RQ2, RQ4	**11/14**	**Provisional include**
10.3389/fpsyt.2025.1703302	fNIRS and deep-learning modelling of emotional face processing in preschool ASD.	RQ1, RQ2, RQ3	**13/14**	**Provisional include**
10.1177/15500594251362402	EEG and eye-tracking evaluation of a game-like iPad emotion-recognition app.	RQ1, RQ4, RQ5	**13/14**	**Provisional include**

Note. The studies listed in this table are reported as provisional supplemental evidence. They are not included in the core PRISMA corpus, the Figure 1 study-selection flow, or the RQ-specific evidence-map frequencies. Their inclusion in a future expanded synthesis requires completion of the same full-text eligibility screening, quality appraisal, and structured data-extraction procedure for all remaining supplemental candidates.

Interpretive use of supplemental evidence

The supplemental source audit preserves the conceptual relevance of multimodal sensing, speech, physiological signals, gesture, and child-specific interaction design without altering RQ1 frequencies. The selected contextual references are used only to interpret why face-based recognition may benefit from complementary modalities and why synchronization, consent, and data minimization require attention [33], [36], [37], [38], [40], [43]-[46]. For RQ2 and RQ3, the contextual literature motivates a cautious, child-centered interpretation of model and dataset findings. Studies addressing computer vision in autism, real-time IoT systems, child-oriented multimodal data, and youth affective-computing protocols suggest that performance transfer from adult datasets should not be assumed without age-aware validation and context-sensitive evaluation

ISI Journal of Information Systems and Informatics
Published By Asosiasi Doktor Sistem Informasi Indonesia

[35], [38]-[40], [48]. These references inform the discussion but do not contribute to the quantitative counts reported in Figures 3 and 4. For RQ4 and RQ5, the contextual literature supports the principle that emotion inference should lead to proportionate, privacy-sensitive action rather than automated diagnosis. Parent-assisted serious games, digital gaming interventions, and privacy-preserving engagement or meltdown detection motivate a design in which alert escalation is accompanied by caregiver control, uncertainty communication, and data minimization [34], [41], [47]. These references are interpretive supports rather than additional included studies.

Additional IEEE Xplore supplemental candidates

Two additional IEEE Xplore records were identified as high-priority supplemental candidates during source verification [53], [54]. These records were not retrieved into the fixed core corpus through the original five-block conjunctive search and therefore remain supplemental sensitivity candidates. They are not incorporated into the core PRISMA corpus, RQ-specific evidence maps, or quality-appraisal summary of the 32 core studies.