AI-based Early Dyslexia Detection: A Systematic Literature Review

Daniel Waardenburg, Clara Maathuis, Ebrahim Rahimi and Petru Curseu
Open University of the Netherlands, Heerlen, The Netherlands

daniel.waardenburg@ou.nl
clara.maathuis@ou.nl
ebrahim.rahimi@ou.nl
petru.curseu@ou.nl

Abstract: AI (Artificial Intelligence) represents a promising tool for the early dyslexia detection as it enables a more timely and targeted intervention in primary education. Despite the rapid research expansion in this domain, its foundations remain uneven since various studies are situated in technical or highly controlled research contexts, with insufficient attention to the educational, linguistic, and organizational dimensions in which such systems need to function in school practice. To address this knowledge gap, this research presents a systematic literature review of AI-based early dyslexia detection and provides a responsible AI governance-oriented analysis to identify the key conditions required for its responsible and usable implementation for children from age four through primary education. Drawing on 61 studies published between 2013 and 2025, the review shows that although technical performance improved across motor, linguistic, and neurocognitive domains, the evidence base remains methodologically fragmented since many findings are derived from small, retrospective, or otherwise constrained datasets that limit their relevance for authentic classroom settings. Additionally, different ethical challenges exist concerning generalizability, systemic bias, transparency, security, and practical usefulness. At the same time, it is observed that linguistic and cultural diversity are addressed inconsistently and bias is frequently acknowledged but rarely translated into concrete design choices or evaluation procedures which makes fairness claims to be provisional. A similar pattern is found in relation to ethical governance: although widely discussed, transparency, accountability, human-in-the-loop oversight, and responsibility for decision-making remain insufficiently specified once systems move beyond experimental settings. At the same time, the AI generated materials are experienced as usable when teacher control, curriculum alignment, and accessible formatting for learners with dyslexia are ensured.

Keywords: Dyslexia, Early dyslexia detection, Neurodivergence, Artificial intelligence, AI based dyslexia detection

1. Introduction

Understanding when dyslexia first becomes observable remains a complex question in both educational and cognitive research domains. Rather than constituting a discrete deficit in reading, dyslexia is more accurately characterized as a multidimensional learning difficulty involving impairments in reading fluency, word recognition, spelling, and broader language-related processing (Ahire et al., 2023). This broader perspective is important since early literacy difficulties rarely can be resolved spontaneously or in a fast tempo. They tend to persist over time, with well-documented consequences for academic achievement, socio-emotional development, and the formation of self-efficacy during the early school years and beyond (Snowling, 2013). Nevertheless, formal identification practices are initiated after literacy problems become clearly evident in familiar/school contexts, thereby constraining opportunities for timely and preventive intervention (Daniel et al., 2024; Snowling, 2013). As a result, educational responses often remain reactive and focus on remediation instead of early risk detection.

At the same time, the persistence of delayed identification cannot be attributed to a lack of observable early indicators. A growing body of empirical studies shows that differences in phonological processing, auditory temporal sensitivity, visual attention, and related developmental mechanisms can be detected prior to or at the very onset of reading educational instructions (Christoforou et al., 2021; Coenen et al., 2024; Ding et al., 2025). However, the existence of such early signals does not imply their immediate applicability in educational practice. Along these lines, translating research-based indicators into tools that are reliable, interpretable, and feasible within real-world school settings remains a significant challenge. These settings deal with challenges like limited time, uneven distribution of specialist expertise, and fragmented observational input across teachers, support staff, and caregivers (Daniel et al., 2024).

Within this context, AI represents a promising direction for advancing early dyslexia identification efforts. As Yap et al., (2025) argues, AI-based solutions enable earlier, more systematic, and scalable detection of dyslexia-related risks which allows specialists to develop and apply tailored methods to support children's development. This is reflected by studies that consider a diverse methodological landscape of AI technologies such as handwriting-based models, speech and language analysis, eye-tracking techniques, electroencephalography (EEG)-based modelling, and multimodal systems that integrate multiple data sources

(Gull & Khan, 2023; Parmar & Paunwala, 2023; Raatikainen et al., 2021; Swami & K M, 2025; Tiwari et al., 2025). Nonetheless, as Johnson (2025), Khosravi et al., (2022) and Khreisat et al., (2024) argue, educational readiness and governance can be achieved by assuring that the implementation of AI solutions is conducted in a responsible way by accounting and respecting aspects like transparency, security, fairness, and reliability. This allows both the children and relevant educational stakeholders to properly use, trust, and rely on the support that these solutions provide (Johnson, 2025; Nazaretsky et al., 2022).

To this end, this research aims to synthesise the current evidence base on AI-based early dyslexia detection for children from age four through primary education and to identify the conditions under which such systems could be implemented in a responsible and educationally usable way in everyday school practice. On this behalf, four research questions (RQs) are defined and a systematic literature review is performed on academic studies conducted between 2013 and 2025 using the PRISMA methodology.

The remainder of this article is structured as follows. Section 2 presents the background and research studies conducted in this domain. Section 3 discusses the methodological approach followed to achieve the aim of this research. Section 4 provides on overview of the core results obtained in this process. And Section 5 reflects on concluding remarks and future research perspectives in this domain.

2. Research Background and Related Studies

AI is increasingly being explored as a way to strengthen the early dyslexia detection in educational settings. This development responds to a long standing practical difficulty. Although early identification was recognized as important, many children are still noticed once reading difficulties already become visible in school performance (Daniel et al., 2024; Snowling, 2013). Along these lines, the rise of AI-based screening reflects a broader move away from delayed recognition and towards earlier, more structured forms of identification, especially in the first years of schooling (Ahire et al., 2023; Alkhurayyif & Sait, 2024; Kurniawan & Tiaharyadini, 2024; Yap et al., 2025).

Ahire et al. (2023), Alkhurayyif and Sait (2024), Gull and Khan (2023), and Yap et al. (2025) discuss the variety of AI paradigms such as the ones based on machine learning and deep learning methods and stress out that such solutions are built from a technical perspective instead of a literacy assessment one. Such applications are dedicated to handwriting analysis, gamified tasks, eye movement data, electroencephalography (EEG) related measures, psychometric systems, and multimodal combinations of behavioural, neurophysiological, and linguistic markers (Aldehim et al., 2024; Alkhurayyif & Sait, 2024; Alluhaidan et al., 2024; Christoforou et al., 2021; Coenen et al., 2024; Hu, 2024; Parmar & Paunwala, 2023; Raatikainen et al., 2021; Rauschenberger et al., 2025; Rello et al., 2020; Swami & K M, 2025; Tiwari et al., 2025; Usman et al., 2021; Zahia et al., 2020; Zhong et al., 2023). This broadening of modalities suggests that early dyslexia detection is now treated as a pattern recognition challenge in which several indicators support screening decisions together instead of relying on just a simple detection test. The themes identified across the related studies are further presented.

The first theme concerns the diversification of data sources. Handwriting-based studies show that visual and motor traces can serve as meaningful inputs for dyslexia related classification (Aldehim et al., 2024; Alkhurayyif & Sait, 2024). Game-based solutions show that AI can support earlier and more scalable forms of identification in potentially more child friendly contexts (Rauschenberger et al., 2025; Rello et al., 2020). Concurrently, Coenen et al. (2024) show the growing role of eye tracking in dyslexia related assessment, while Parmar and Paunwala (2023) indicate that EEG data may also provide useful discriminatory information. Tiwari et al. (2025) further emphasize the potential of multimodal combinations. These studies show that that the technical diversity is proportional with the interpretation complexity because performance under controlled conditions does not automatically translate into educational usability in real school settings.

The second theme concerns the uneven methodological foundations of the solutions proposed in this domain. This exists due to the fact that while various studies report promising outcomes, they are often based on small samples, retrospective datasets, narrowly defined linguistic contexts, or highly specific task environments (Ahire et al., 2023; Alkhurayyif & Sait, 2024; Gull & Khan, 2023; Usman et al., 2021; Yap et al., 2025). This does not invalidate the findings, it limits the confidence with which such results can be transferred to everyday educational practice. In primary education, screening tools need to function across heterogeneous groups of children instead of within tightly controlled research settings. Hence, the challenge is the reliability and robustness of these solutions in classroom which is dependent of factors such as psychometric limitations as Verpalen et al., (2018) stresses. This translates to the importance of the principles and choices embedded in

their design and evaluation phases which need to include the needs and constraints of all the stakeholders involved as well as the applicable institutional constraints.

The third theme concerns linguistic and contextual variation. Phonological processing remains central in dyslexia research, yet current AI-based detection extends beyond phonology alone (Christoforou et al., 2021; Coenen et al., 2024; Ding et al., 2025; El Hmimdi & Kapoula, 2025; Luke et al., 2024; Nguyen et al., 2025; Parmar & Paunwala, 2023; Raatikainen et al., 2021; Swami & K M, 2025; Toki, 2024). For instance, cross linguistic differences and orthographic depth matter and dyslexia related patterns do not emerge in identical ways across languages and learner groups. Along these lines, Ding et al. (2025), Verpalen et al (2018)., and Zhong et al. (2023) emphasize that orthographic or cultural settings from one context cannot simply be assumed to transfer well to another. This is important for any attempt to move towards practical school implementation because it directly affects how claims of effectiveness should be interpreted by teachers, schools, and policymakers.

Additionally, Johnson (2025), Khosravi et al. (2022), and Khreisat et al. (2024) consider that responsible AI in education requires more than predictive performance. It also requires explainability, fairness, accountability, and clarity about the role of human judgement in acting on model outputs. Nazaretsky et al. (2022) further show that teacher trust also depends on whether the systems are understandable and usable in practice. Moreover, (Johnson, 2025) considers that fairness is a major challenge still inconsistently operationalized. Therefore, understanding how and under which conditions AI is being designed and evaluated in a responsible way for early dyslexia detection in educational settings needs to be first understood. This represents the knowledge gap that this article aims to tackle.

3. Research Methodology

This research aims to study how AI is being designed and used in the context of early dyslexia detection and how such use can be aligned with responsible AI governance in educational settings. To this end, this research seeks to identify and synthesize evidence from existing AI-based tools, relevant frameworks, and policy-oriented approaches to clarify what is currently known about their intended purpose, operational scope, methodological strengths, practical limitations, and suitability for classroom use. At the same time, the study aims to understand how these approaches address core responsible AI governance concerns such as transparency, fairness, accountability, privacy, robustness, and the extent to which they respond to the realities of educational practice. Through this effort, the review is designed to provide a structured basis for the development of responsible AI systems for early dyslexia detection.

To achieve this aim, a systematic literature review is conducted following the guidelines of the PRISMA research methodology dedicated to answering the following four RQs are formulated as follows:

- RQ1: What can be learned from existing AI systems for early dyslexia detection and from existing frameworks and policies regarding their scope, strengths, limitations, governance requirements, and classroom relevance?
- RQ2: How are the needs and expectations of the teachers, parents, and policymakers translated into design requirements and principles in a responsible AI governance framework?
- RQ3: How can such a responsible AI governance framework be structured so that the defined requirements and principles translate into clear and usable guidelines for various stakeholders?
- RQ4: How can the effectiveness of such a framework be assessed in real world classroom contexts, and what lessons can be drawn from its implementation for improving the responsible use of AI in education?

In this review process (see Figure 1a), five scientific databases were considered: Web of Science, Scopus, IEEE Xplore, ACM Digital Library, and PubMed. The studies collected were conducted between 01.01.2013 and 31.12.2025 and had to be written in English. The search query included keywords such as AI, Generative AI, Responsible AI, dyslexia detection, and early dyslexia detection. In the initial phase, 197 records were identified through scientific database searching. Before formal screening, 79 records were removed because they were duplicates or irrelevant on the basis of title level inspection. This left 118 studies for title and abstract screening. In the screening phase, 21 studies were excluded because they did not focus on early dyslexia detection, were not sufficiently related to AI supported identification, or showed no relevance to the review scope. This resulted in 97 reports sought for retrieval. In the eligibility phase, 31 reports could not be retrieved given reasons such as availability given institutional subscriptions, publisher platforms, or accessible author correspondence. The remaining 66 full studies were assessed in detail, after which 5 studies were

excluded because their relevance to AI aided early dyslexia detection was limited. The final corpus therefore consisted of 61 included studies distributed in various years as illustrated in Figure 2.

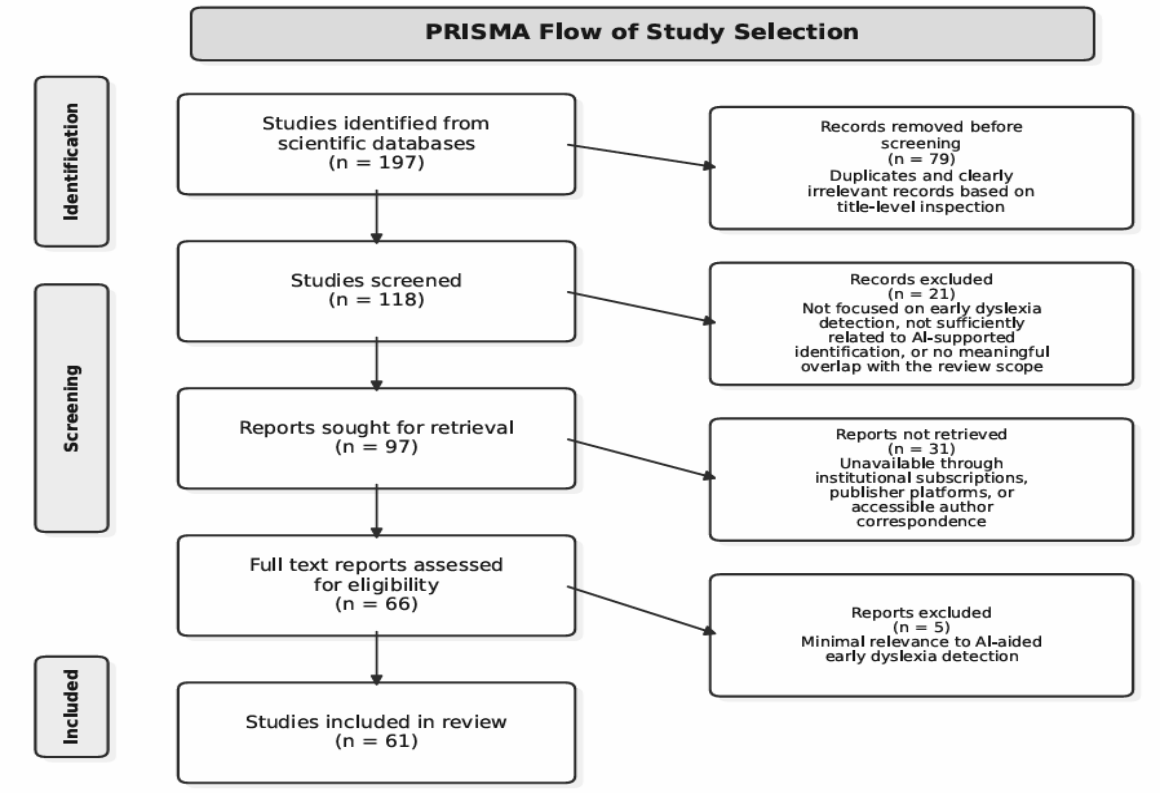


Figure 1a: The PRISMA flow diagram tailored to this review.

The year distribution suggests that research on this topic has grown in recent years, with only isolated publications before 2020 and a clear acceleration from 2020 onward. The steady rise through 2022 and 2023, followed by the strong peak in 2024 and the similarly high count in 2025 indicates that the field is currently gaining substantial momentum. This pattern likely reflects increasing academic and practical interest, growing data availability, and wider use of AI methods for studying and supporting dyslexia-related detection, assessment, and intervention. Hence, the included studies form a heterogeneous evidence base comprising predictive and diagnostic modelling studies, systematic and scoping reviews, conceptual and governance-oriented papers, application focused studies, and other empirical contributions. To analyse this diverse body of work, the review applied shared analytical lenses across the included studies, namely data modality, target population, validation strategy, educational usability, and governance relevance. A detailed classification of the 61 included studies by sample size, age range, language or orthography, and validation methodology is provided in Figure 1b.

Article	Sample size	Age range	Language	Validation
Ahire et al. (2023)	n/a (review aggregates multiple primary studies with differing sample sizes; no sing…	Primary EEG studies typically involve school-age children (≈7–12 years) with diagnos…	NR	Review
Aldehim et al. (2024)	49,212 handwriting images (36,912 training; 4,100 validation; 8,200 testing)	Age of writers not reported; NIST 19 typically contains adult and older-student writ…	NR	NR
AlGhamdi (2022)	N = 3,644.	Dataset covers ages 7–17 years.	NR	CV + TTS
Alkhurayyif (2023)	123,554 handwritten images (train: 86,115; test: 37,439; normal + reversal)	Intended for early-stage school children with handwriting samples; dataset mixes pri…	NR	NR
Alkhurayyif et al. (2024)	n/a (review of 39 studies with heterogeneous sample sizes)	Children and adults, mostly school-age and young adults; very limited direct evidenc…	NR	Review
Alluhaidan et al. (2024)	3,644 participants, ages 7–17, of whom 392 have diagnosed dyslexia (re-used PLOS ONE…	Directly covers ages 7–17; sub-cohort A6 includes 7–8-year-olds using age-adapted ta…	Spanish	NR
Christoforou et al. (2021)	60 children (30 DYS, 30 controls) (PubMed)	7.6–12.1 years (M ≈ 9.8); above AERAIFD target 4–8.	NR	NR
Coenen et al. (2024)	N = 1,566 children across 20 primary studies (549 experimental, 618 control, others…	The review targets school-aged children 6–12 years; 17 of 20 studies focus on a core…	English; Cross-linguistic	Systematic review
Cogan et al. (2025)	84 participants; 230,412 eye-tracking records from the Provo Corpus	Age ~18–30 years (university cohort; exact range not reported).	English	NR
da Silva et al. (2020)	N=60 (30 dyslexia, 30 control) (Frontiers)	Ages 7–12 (mean ≈9.7 years).	Portuguese	NR
Dabaghi et al. (2024)	172 included papers	Mixed educational levels; no single participant age range reported	English-language publications	Review with independent screening and content classification; no primary model validation
Daniel et al. (2025)	274 professionals	School-aged range from reception/pre-school (~4–5 years) and Key Stage 1 (years 1–2;…	English	Survey
Ding et al. (2025)	1196 screened; 263 in analytic sample (128 DD, 135 controls)	Focuses on literate primary-school children, not 4–8-year-old pre-readers; however,…	Chinese	CV
Eitel-Porter (2021)	n/a (no quantitative or qualitative dataset).	Focus is on adult stakeholders (executives, employees, customers, regulators); child…	NR	NR
El Hmimdi et al. (2025)	355 individuals for saccade task; 454 individuals for vergence task from Ora22 datas…	Ora22 subset includes individuals with median age 10–12 years in all three groups; a…	NR	NR
Ergasheva (2025)	n/a (no empirical dataset; narrative discussion only) (Academic Publishers)	Refers broadly to “early detection” and early school grades; implicitly targets firs…	NR	NR
Gasmi et al. (2024)	3,644 participants (392 with dyslexia, 10.7%)	7–17 years; roughly primary to secondary school; includes early readers (7–9) but no…	NR	NR
Goldman et al. (2024)	n/a (no empirical sample; guidance article).	Student age not specified; examples likely span K–12.	NR	NR
Goodman et al. (2022)	N=19	Participants were adults aged 18–54 years (one 18–24, seven 25–34, eleven 35–54).	English	Interview / design
Gull et al. (2023)	n/a (review across 30–40 studies; exact count not accessible) (SpringerLink)	Typically 7–16 years; some studies include pre-readers, but coverage of 4–6 years is…	NR	Systematic review
Harkins-Brown et al. (2025)	NR	Covers young children, school-age students and young adults; notes AI in early child…	NR	Systematic review
Hu (2024)	n/a (conceptual working paper; no empirical sample).	School-age children and adolescents with learning difficulties; the paper cites epid…	NR	NR
Iyer et al. (2023)	n/a (no primary empirical sample; synthesises multiple prior studies)	The chapter generally refers to “children” in primary education and special educatio…	NR	NR
Johnson (2025)	n/a (no empirical sample; conceptual synthesis).	Lifelong learning from schooling to workforce.	NR	NR
Khosravi et al. (2022)	n/a (framework paper drawing on several prior empirical studies; no single consolida…	Primarily adolescents and adults (secondary school, university, and professional tra…	NR	NR
Khreisat et al. (2024)	n/a (number of included studies not explicitly reported; 43 references listed)	The reviewed literature spans school and higher-education settings without specifyin…	NR	Systematic review
Kurniawan et al. (2024)	100 children	Exact ages not stated; likely about 6–12 in Indonesian primary education (inferred f…	NR	Scoping review
Leppänen et al. (2019)	N = 391 with usable EEG data: 195 dyslexic readers and 196 controls	8–12 years; mean age 9.97 years	Finnish, Hungarian, German and French	Four-country EEG replication study

Luke et al. (2024)	95 participants (39 dyslexia, 56 control)	Mean age ~22 years; adult readers only; no children or 4–8-year-olds included; any i…	English	NR
Martinez-Murcia et al. (2020)	n/a (sample size not reported in accessible abstract; likely school-based cohort) (a…	The model targets children who can complete phonological tests at about 5 years of a…	NR	NR
Nazaretsky et al. (2022)	≈39 teachers (N≈36–39 across analyses; K-12 science) (ResearchGate)	Teachers of students roughly 12–18 years; indirectly informs how tools for children…	NR	Survey
Nguyen et al. (2025)	n/a (sample size not available in open metadata)	Likely adult or adolescent readers; no explicit child data.	NR	NR
Pamungkas et al. (2025)	31 studies; individual samples from ≈30 to >3,600 participants	Mainly 7–17 years; some studies include younger children (down to 4) and adults; evi…	NR	Systematic review
Parmar et al. (2023)	N = 53 (29 dyslexic, 24 controls)	Dataset: 7–12-year-olds; implementation goal: early-detection system for children 8–…	NR	CV
Patel et al. (2025)	Number of included studies not clearly specified in accessible metadata; user notes…	Focus appears to be on school-age learners (approximately 6–18 years), with "early d…	NR	NR
Perera et al. (2018)	N/A; critical review of multiple EEG studies with different sample sizes	Children and adults; reviewed child samples ranged approximately from 2–16 years	Not systematically reported	Critical review
Prasad (2022)	n/a (sample details not accessible)	Likely primary-school and possibly middle-school students; no evidence for inclusion…	Hindi	NR
Raatikainen et al. (2021)	N = 161 (30 labelled with reading disorder)	Youngsters with an average age of 12.5 years (approx.	Finnish	CV
Ram et al. (2025)	37 students (survey) + 3 SSWS staff (interviews).	Target users are university students, typically aged 18–25, with some mature student…	NR	Survey
Rauschenberger et al. (2025)	137 children (51 with dyslexia; 86 controls)	Approx.	NR	NR
Rello et al. (2020)	N=3,644 (392 with dyslexia, 3,252 without) (PLOS)	Participants aged 7–17 years; exercises are designed for children seven or older, wi…	Spanish	CV
Robaa et al. (2024)	810,000 character images overall; tri-class subset 151,750 (40,000 normal, 46,750 re…	Age is not reported; NIST SD19 and Kaggle likely contain mainly adolescents/adults,…	NR	CV
Rokade et al. (2024)	n/a (no empirical evaluation reported)	Age not stated; narrative suggests early school-age to adolescence; envisioned as "e…	NR	NR
Sbiaa et al. (2024)	n/a (sample size not reported in abstract or secondary summaries; full text not acce…	Participants are described as "young learners" and "children"; accessible summaries…	NR	TTS
Shahini et al. (2025)	84 primary studies (2018–2024)	Focus on paediatric populations roughly 4–18 years; several tools for early childhoo…	NR	Systematic review
Shaywitz et al. (2021)	N/A (review article; no primary participant sample)	Children from first grade onward; no single empirical age range reported		Review article; no primary validation
Sigurdardottir et al. (2021)	N = 60 adults: 24 dyslexic readers and 36 typical readers	19–51 years; mean age 25 years	Icelandic; all participants were native Icelandic speakers	Cross-sectional behavioural study using face-matching and reading tasks; between-
Snowling (2013)	n/a (narrative review; no single empirical sample)	Focus on preschool and early primary years in the UK system: pre-reading and emergen…	NR	NR
Song et al. (2024)	n/a (framework paper; no single pooled empirical sample)	Core focus is on K–12 broadly, with illustrative application in a middle-school AI c…	NR	Review
Swami et al. (2025)	1000 participants (500 dyslexic, 500 non-dyslexic) inferred from confusion matrix (T…	Age is not reported.	NR	Review
Tiwari et al. (2025)	70 children in ETDD70 eye-tracking dataset (aged 9–10); additional pilot users for p…	Empirically validated data come from 9–10-year-old Czech schoolchildren; the framewo…	Multilingual; Cross-lin	NR
Toki (2024)	21 primary studies; total N across studies not consistently reported (systematic rev…	The review includes children 0–18 by inclusion criteria, but most primary studies te…	English	Systematic review
Usman et al. (2021)	22 included ML studies (primary samples range approx.	Predominantly school-age children (≈6–12 years) and adolescents; some studies includ…	NR	Review
Vajs et al. (2023)	30 children; 15 with dyslexia, 15 controls (MDPI)	Ages 7–13; for 4–8 year-olds, gaze-complexity measures could be reused with much sim…	NR	NR
Velmurugan (2023)	n/a (review of multiple primary studies; no single pooled sample)	Reviewed datasets span from preschool to adulthood: a German longitudinal cohort rec…	NR	Review
Verpalen et al. (2018)	~800–850 children (exact N varies by analysis) (PMC)	Children in grades roughly equivalent to 3–6 (approx.	Dutch; Multilingual	NR
Yap et al. (2025)	84 included studies	Mixed ages across included studies	Multilingual	Scoping review
Zahia et al. (2020)	55 children (19 dyslexic, 17 monocular vision, 19 controls); 165 fMRI activation vol…	Children initially 66, 55 retained, aged 9–12 years (mean ≈10–10.5) from paediatric…	Spanish	CV + TTS
Zainuddin et al. (2022)	N=54 (18 poor dyslexic, 18 capable dyslexic, 18 controls)	Children; exact age NR	NR	10-fold CV
Zhong et al. (2023)	N=207	6–8 years	Chinese	CV
Zingoni et al. (2024)	1,217 questionnaire responses for model development; 43 students in real-world pilot	Adults only: inclusion criterion "being more than 18 years old" and in/near universi…	NR	CV

Legend: CV = cross-validation; TTS = train-test split; EV = external validation; NR = not reported.

Figure 1b: Classification of the 61 included studies by sample size, age range, language or orthography, and validation methodology

This allowed identifying both technical progress and recurring limitations, particularly in relation to Responsible AI principles and aspects such as generalizability, representativeness, fairness, transparency, security, and practical usability in real primary school settings.

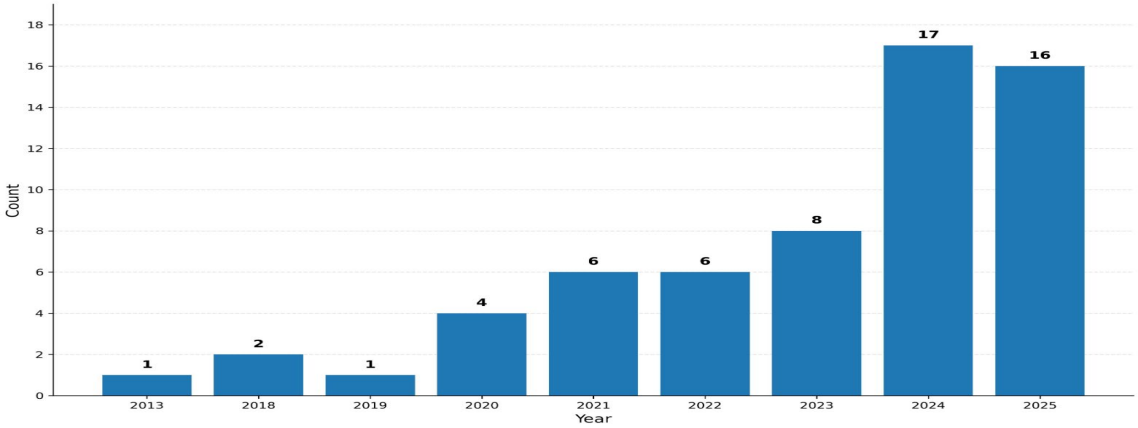


Figure 2: Distribution per year for the studies analysed

4. Results

This section presents the results of the study in relation to the four research questions formulated in this article. These results are complemented by the dominant themes depicted in Figure 3 from the reviewed studies. The review covered publications from 2013 to 2025. Figure 2 shows one included study from 2013, followed by the remaining included studies published between 2018 and 2025. No included studies were recorded for the period 2014–2017.

The prominence of terms such as dyslexia, learning, children, reading, detection, artificial, data, accuracy, performance, and eye-tracking shows that (i) the core of developing and using AI-based systems in this context is mainly dedicated to reading and learning purposes. The word cloud is used as a descriptive illustration of recurring themes, while the conclusion that the majority of the studies are centred on technical dimensions is based on the systematic classification of the reviewed studies.

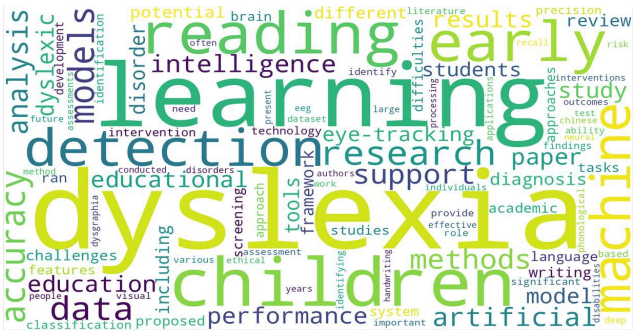


Figure 3: Dominant topics based on abstract analysis.

4.1 Existing AI Systems

Existing studies show that dyslexia-related screening now includes handwriting-based models, game-based screening, eye-tracking approaches, electroencephalography (EEG)-related modelling, psychometric systems, and multimodal combinations of behavioral, linguistic, and neurophysiological markers (Ahire et al., 2023; Alkhurayyif & Sait, 2024). This broadening from single detection to multimodal-based detection matters because it reveals that early dyslexia-related risk is increasingly approached as a pattern-recognition problem in which several indicators may together support screening decisions, rather than as a single-test classification problem (Gull & Khan, 2023; Yap et al., 2025).

With recent technological developments, the performance of the AI models improved substantially across a number of modalities, especially in controlled benchmark settings. Accordingly, handwriting-based models report high classification scores, e.g., accuracy levels above 95 percent in some studies (Aldehim et al., 2024; Alkhurayyif & Sait, 2023). In gamified and online settings, practical potential for lower-burden early screening is achieved (Alluhaidan et al., 2024; Rello et al., 2020). In eye-tracking contexts, consistent discriminatory

patterns such as atypical fixation behavior, altered saccades, or other gaze-based indicators are observed and are able to support non-invasive detection pathways (Coenen et al., 2024; Raatikainen et al., 2021). Moreover, multimodal studies strengthen this pattern by suggesting that combined data streams may improve robustness beyond single-modality tools (Swami & K M, 2025; Tiwari et al., 2025).

In terms of data used, various studies rely on small samples, retrospective datasets, single-language settings, or narrowly defined task environments (Ahire et al., 2023; Alkhurayyif & Sait, 2024), and some systems were developed and tested on only one dataset, without external validation or prospective classroom testing (Aldehim et al., 2024; AlGhamdi, 2022). In other cases, demographic information was incomplete or dyslexia labels were inherited from earlier datasets rather than revalidated in the new study context (Aldehim et al., 2024; Alkhurayyif & Sait, 2023).

A closely related finding concerns age range and school relevance. Although this review is positioned from age four through primary education, much of the direct evidence still begins later, often at age seven or above (AlGhamdi, 2022; Alluhaidan et al., 2024; Parmar & Paunwala, 2023; Rello et al., 2020). This does not invalidate the field, but it does show that the strongest current evidence concerns early and middle primary school more often than true pre-reading or preschool screening. In that sense, the literature supports early identification at a conceptual level, but lacks consistency at the youngest end of the target range. The review therefore suggests that claims about “early detection” should be used with precision. In many cases, the tools are early relative to formal diagnosis and not necessarily early in the sense of pre-reading identification.

Additionally, classroom relevance remains underdeveloped. Existing AI tools are usually designed as technical classifiers, decision-support, or prototype screening systems instead of school-integrated interventions (Alkhurayyif & Sait, 2024; Gull & Khan, 2023). Their outputs are often framed in terms of metrics like accuracy, recall, or area under the curve, while questions about interpretability, teacher-facing usability, referral logic, and implementation workflow remain unspecified (Ahire et al., 2023; Usman et al., 2021). This gap becomes even more visible when the technical literature is read alongside governance-oriented studies. Responsible AI studies in education stress that screening systems need more than predictive performance. They also require transparency, explainability, accountability, human oversight, and alignment with educational purpose (Johnson, 2025; Khosravi et al., 2022; Khreisat et al., 2024). Therefore, the current tool landscape is technically promising and increasingly multimodal while providing only partial clarity on when those tools are sufficiently robust, interpretable, and context-sensitive for real school use.

4.2 Stakeholders’ Needs Assessment

The first design requirement is that AI-supported detection needs to remain transparent, understandable, and reliable for professionals in practice. Studies on teacher trust show that educators do not respond positively to AI simply because it performs well. Trust depends on whether outputs are interpretable, whether teachers retain agency, and whether the system is seen as supporting rather than replacing professional judgement (Nazaretsky et al., 2022). Related governance studies reinforce this by stressing the need for explainability, documentation, and clear responsibility structures around automated recommendations (Johnson, 2025; Khosravi et al., 2022; Khreisat et al., 2024).

The second design requirement concerns workload, accessibility, and practical fit. Several studies suggest that AI becomes educationally useful only when it reduces rather than increases professional burden (Goldman et al., 2024; Nazaretsky et al., 2022). This matters in dyslexia-related screening because schools already operate under time pressure, uneven expertise, and fragmented observation practices. Game-based screening and digitally mediated tasks are relevant here because they suggest more scalable and potentially child- friendly routes into early identification (Alluhaidan et al., 2024; Rauschenberger et al., 2025; Rello et al., 2020). At the same time, child-appropriate design, curriculum relevance, manageable testing time, and accessible feedback for professionals need to be considered.

The third design requirement concerns fairness, inclusion, and linguistic diversity. This theme appears consistently across the literature while being partially operationalized in the technical studies (Johnson, 2025). Furthermore, cross-linguistic studies show that dyslexia-related patterns do not emerge in identical ways across languages and learner groups (Ding et al., 2025; Verpalen et al., 2018; Zhong et al., 2023). Hence, implementation requires subgroup awareness, representative training data, and explicit caution around transfer claims between orthographies, cultural settings, or multilingual populations. In this sense, the review suggests that fairness should be treated as a concrete design requirement.

The fourth design requirement concerns decision pathways and the role of screening itself. Along these lines, AI-based systems should not be framed as autonomous diagnostic systems since they should have a supportive role for risk signaling, structured observation, and better informed referral or follow-up decisions (Johnson, 2025; Khreisat et al., 2024). This distinction matters for teachers, parents, and policymakers. Parents need systems that do not stigmatize children through premature or opaque classification and policymakers need systems that fit governance rules, data protection responsibilities, and equitable access obligations. Therefore, the needs of these stakeholders need to be translated into a common set of requirements that concerns aspects such as transparency, human oversight, data representation and use, referral boundaries about what the tool can and cannot do.

4.3 Towards a Responsible AI Governance Framework

Based on the review conducted in this research, a Responsible AI governance framework needs to have a clear scope and goal where AI is positioned as a tool for screening and decision-support instead of a stand-alone diagnostic instrument. Accordingly, a set of design principles is further presented. The first design principle is justified by both the current technical limitations of available models and the broader governance literature on responsible educational AI (Johnson, 2025; Khreisat et al., 2024). The second design principle concerns the data representativeness and quality since the reviewed studies indicate that promising model performance is often achieved on relatively narrow datasets, making external validity a primary design requirement rather than a secondary consideration (Ahire et al., 2023; Alkhurayyif & Sait, 2024; Usman et al., 2021). Closely related to this is the third design principle which concerns explainability and human oversight since the teachers trust, educational usability, and ethical legitimacy depend on AI outputs that can be interpreted, questioned, and contextualized by professionals rather than passively accepted (Khosravi et al., 2022; Nazaretsky et al., 2022).

The fourth design principle refers to considering stakeholders' usability to ensure that tools remain low-burden, intelligible, and compatible with existing educational workflows. This includes fairness monitoring through subgroup analysis, transparent reporting of system constraints, and particular caution in multilingual or cross-orthographic contexts where model performance and equity risks may differ substantially (Ding et al., 2025). Finally, the fifth design principle refers to including clear governance documentation that integrates accountability structures, security and privacy protection, and continuous monitoring mechanisms to support responsible implementation over time (Johnson, 2025; Khreisat et al., 2024). Hence, these design principles provide a scientifically grounded foundation that serves as the basis for an initial governance architecture for Responsible AI in early dyslexia detection.

4.4 Implications for Framework Evaluation

Various of the studies analysed conduct a technical evaluation limited to offline validation, benchmark comparison, or laboratory-level testing instead of examining the performance of the systems proposed in authentic educational screening settings (Aldehim et al., 2024; Alluhaidan et al., 2024; Parmar & Paunwala, 2023; Swami & K M, 2025). At the same time, a smaller body of literature considers a broader evaluation perspective. For instance, Nazaretsky et al., (2022) stresses the importance of teacher trust which depends on interpretability and user control. Similarly, studies on teacher workload suggest that the practical value of AI is closely tied to whether it meaningfully reduces time pressure, supports the educational tasks, and assures a safe environment (Goldman et al., 2024).

Hence, while technical evaluation metrics are relevant they are insufficient as stand-alone indicators of success. Future classroom-based evaluation need to therefore also address transferability across school contexts, subgroup fairness, interpretability of outputs, teacher trust, workload effects, referral quality, and governance compliance (Johnson, 2025; Nazaretsky et al., 2022). Hence, this literature review shows that the existing studies and evidence portray a field that is technologically promising but not yet educationally settled in a responsible way. For this reason, the evaluation of these systems needs to go beyond technical evaluation and model comparison to encompass the broader governance and design conditions under which such tools may become responsibly designed and usable in real education contexts.

5. Conclusions

This research examined scientific studies published between 2013 and 2025 on AI-based early dyslexia detection in children from age four through primary education, with a focus on whether such systems can be implemented in a responsible and educationally usable way. The findings indicate a field with strong technical momentum but more limited readiness for real school practice. The reviewed studies show that AI was applied

across a wide range of modalities and applications such as handwriting, gamified tasks, eye tracking, electroencephalography, psychometric systems, and multimodal combinations of behavioural, linguistic, and neurophysiological indicators (Ahire et al., 2023; Alkhurayyif & Sait, 2024; Gull & Khan, 2023; Usman et al., 2021; Yap et al., 2025). This suggests that early dyslexia-related screening is increasingly understood as a multidimensional pattern-recognition problem. At the same time, the review shows that technical promise alone is not sufficient for responsible AI educational use. The evidence remains methodologically uneven and often context-bound while key governance issues such as transparency, fairness, security, accountability, human oversight, and user trust are still insufficiently addressed in the literature (Johnson, 2025; Khosravi et al., 2022; Khreisat et al., 2024; Nazaretsky et al., 2022).

This review is also subject to several limitations. Although the final corpus included 61 studies, 31 of the 97 reports sought for retrieval could not be accessed. Their exclusion may have introduced some availability bias and is therefore acknowledged as a limitation. The included studies also vary substantially in modality, age range, dataset structure, and validation strategy, which limits direct comparison and transferability of conclusions. Moreover, as a systematic literature re view, this study does not validate a governance framework in live educational settings, it identifies the main design requirements and principles that such a framework needs to address. Future research considers three directions. First, conduct a series of expert-based sessions with teachers, specialists, parents, children, and policymakers to better understand their needs, trust, and usability conditions for responsible AI governance implementation. Second, conduct more empirical studies in real school environments with more concrete operationalization of diverse responsible AI principles considering multilingual and cross-orthographic settings. And third, conduct empirical research and a systematic analysis of existing tools and platforms to assess their methodological foundations, fairness, explainability, governance features, and educational fit, thereby supporting the development of a more robust and practice-oriented responsible AI framework for early dyslexia detection.

Ethics Declaration: This study is based on a systematic literature review of previously published academic literature. No human participants were recruited, no intervention was conducted, and no personal data were collected or processed. Therefore, ethical clearance was not required for this research.

AI Declaration: AI based tools were used only for language refinement and grammatical correction. The content, analysis, and conclusions remained entirely the authors' own.

References

Ahire, N., Awale, R.N., Patnaik, S. and Wagh, A. (2023) "A comprehensive review of machine learning approaches for dyslexia diagnosis", Multimedia Tools and Applications, Vol. 82, pp. 13557–13577.

Aldehim, G., Rashid, M., Alluhaidan, A.S., Sakri, S. and Basheer, S. (2024) "Deep Learning for Dyslexia Detection: A Comprehensive CNN Approach with Handwriting Analysis and Benchmark Comparisons", Journal of Disability Research, Vol. 3, Article e20240010.

AlGhamdi, A.S. (2022) "Novel Ensemble Model Recommendation Approach for the Detection of Dyslexia", Children, Vol. 9, No. 9, Article 1337.

Alkhurayyif, Y. and Sait, A.R.W. (2023) "Deep Learning-Based Model for Detecting Dyslexia Using Handwritten Images", Journal of Disability Research, Vol. 2, No. 4, pp. 89–98. doi:10.57197/JDR-2023-0059.

Alkhurayyif, Y. and Sait, A.R.W. (2024) "A review of artificial intelligence-based dyslexia detection techniques", Diagnostics, Vol. 14, No. 21, Article 2362. doi:10.3390/diagnostics14212362.

Alluhaidan, A.S., Rashid, M., Aldehim, G., Basheer, S. and Sakri, S. (2024) "Efficient deep learning model for detecting dyslexia in children from an online gamified test dataset", Journal of Disability Research, Vol. 3, Article e20240099.

Christoforou, C., Fella, A., Leppänen, P.H.T., Georgiou, G.K. and Papadopoulos, T.C. (2021) "Fixation-related potentials in naming speed: A combined EEG and eye-tracking study on children with dyslexia", Clinical Neurophysiology, Vol. 132, No. 11, pp. 2798–2807.

Coenen, L., Grünke, M., Becker-Genschow, S., Schlüter, K., Schulden, M. and Barwasser, A. (2024) "A systematic review of eye-tracking technology in dyslexia diagnosis", Insights into Learning Disabilities, Vol. 21, No. 1, pp. 45–65.

Daniel, J., Clucas, L. and Wang, H.-H. (2024) "Identifying students with dyslexia: Exploration of current assessment methods", Annals of Dyslexia, Vol. 75, pp. 19–41.

Ding, N., Peng, P., Tang, J., Ding, Y. and Zhao, J. (2025) "An investigation of phonological predictors in Chinese developmental dyslexia using a machine learning approach", Reading and Writing, Vol. 38, pp. 2849–2873.

El Hmimdi, A.E. and Kapoula, Z. (2025) "Distinguishing dyslexia, attention deficit, and learning disorders: Insights from AI and eye movements", Bioengineering, Vol. 12, No. 7, Article 737.

Goldman, S.R., Taylor, J., Carreon, A. and Smith, S.J. (2024) "Using AI to Support Special Education Teacher Workload", Journal of Special Education Technology, Vol. 39, No. 3, pp. 434–447.

Gull, T. and Khan, S.M. (2023) "A Systematic Review of Research Dimensions Towards Dyslexia Screening Using Machine Learning", Journal of The Institution of Engineers (India): Series B, Vol. 104, No. 3, pp. 511–522.

Hu, A. (2024) “Developing an AI-Based Psychometric System for Assessing Learning Difficulties and Adaptive System to Overcome: A Qualitative and Conceptual Framework”, non-peer reviewed arXiv preprint, arXiv:2403.06284. doi:10.48550/arXiv.2403.06284.

Johnson, M.S. (2025) Responsible AI for measurement and learning: Principles and practices. Research Report No. RR-25-03. ETS

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S. and Gašević, D. (2022) “Explainable artificial intelligence in education”, Computers and Education: Artificial Intelligence, Vol. 3, Article 100074.

Khreisat, M.N., Khilani, D., Rusho, M.A., Karkkulainen, E.A., Tabuena, A.C. and Diaz Uberas, A. (2024) “Ethical implications of AI integration in educational decision making: Systematic review”, Educational Administration: Theory and Practice, Vol. 30, No. 5, pp. 8521–8527.

Kurniawan, Z. and Tiaharyadini, R. (2024) “Machine Learning Approach for Early Diagnosis of Dyslexia Among Primary School Children: A Scoping Review and Model Development”, Indonesian Journal of Artificial Intelligence and Data Mining, Vol. 7, No. 2, pp. 425–435. doi:10.24014/ijaidm.v7i2.30614.

Luke, S.G., Tolley, C., Gutierrez, A., Smith, C., Brown, T., Woodruff, K. and Ford, O. (2024) “The perceptual span in dyslexic reading and visual search”, Dyslexia, Vol. 30, No. 4, Article e1783.

Nazaretsky, E., Grinberg, N. and Yair, Y. (2022) “Teachers’ trust in AI-powered educational technology and a professional development program to improve it”, British Journal of Educational Technology, Vol. 53, No. 4, pp. 914–931.

Nguyen, Q.-T., Nguyen, H., Tang, Q.-H., Truong, T., Pham, V.-T., Le, L. and Williams-King, D. (2025) “Learning disorder detection using eye tracking: Are large language models better than machine learning?”, in Proceedings of the 2025 ACM Symposium on Eye Tracking Research and Applications (ETRA ’25). New York: ACM.

Parmar, S. and Paunwala, C. (2023) “Early detection of dyslexia based on EEG with novel predictor extraction and selection”, Discover Artificial Intelligence, Vol. 3, Article 33.

Raatikainen, P., Hautala, J., Loberg, O., Kärkkäinen, T., Leppänen, P. and Nieminen, P. (2021) “Detection of developmental dyslexia with machine learning using eye movement data”, Array, Vol. 12, Article 100087.

Rauschenberger, M., Baeza-Yates, R. and Rello, L. (2025) “Screening dyslexia using visual auditory computer games and machine learning”, IEEE Access, Vol. 13, pp. 29541–29553.

Rello, L., Baeza-Yates, R., Ali, A., Bigham, J.P. and Serra, M. (2020) “Predicting risk of dyslexia with an online gamified test”, PLOS ONE, Vol. 15, No. 12, Article e0241687.

Snowling, M.J. (2013) “Early identification and interventions for dyslexia: A contemporary view”, Journal of Research in Special Educational Needs, Vol. 13, No. 1, pp. 7–14.

Swami, G. and K M, Y. (2025) “Early Detection of Dyslexia Using Multimodal Analysis of Behavioral, Neurophysiological and Linguistic Markers”, in Proceedings of the 3rd International Conference on Futuristic Technology (INCOFT 2025). Vol. 3. Setúbal: SCITEPRESS, pp. 338–345. doi:10.5220/0013615600004664.

Tiwari, V., Agarwal, O., Sharma, M., Sahu, R., Babar, R., Geddam, R., Awais, M. and Ghayvat, H. (2025) “Akshar Mitra: a multimodal integrated framework for early dyslexia detection”, Frontiers in Digital Health, Vol. 7, Article 1726307. doi:10.3389/fdgth.2025.1726307.

Toki, E.I. (2024) “Using eye-tracking to assess dyslexia: A systematic review of emerging evidence”, Education Sciences, Vol. 14, No. 11, Article 1256.

Usman, O.L., Muniyandi, R.C., Omar, K. and Mohamad, M. (2021) “Advance machine learning methods for dyslexia biomarker detection: A review of implementation details and challenges”, IEEE Access, Vol. 9, pp. 36879–36896.

Verpalen, A., Van de Vijver, F. and Backus, A. (2018) “Bias in dyslexia screening in a Dutch multicultural population”, Annals of Dyslexia, Vol. 68, pp. 43–68.

Yap, J.R., Aruthanan, T. and Chin, M. (2025) “Artificial intelligence in dyslexia research and education: A scoping review”, IEEE Access, Vol. 13, pp. 7123–7134.

Zahia, S., Garcia-Zapirain, B., Saralegui, I. and Fernandez-Ruanova, B. (2020) “Dyslexia detection using 3D convolutional neural networks and functional magnetic resonance imaging”, Computer Methods and Programs in Biomedicine, Vol. 197, Article 105726.

Zhong, S., Song, S., Tang, T., Nie, F., Zhou, X., Zhao, Y., Zhao, Y., Sin, K.F. and Chan, S.-H.G. (2023) “DYPA: A machine learning dyslexia prescreening mobile application for Chinese children”, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, Vol. 7, No. 3, Article 143.