Applied Sciences, Computing and Enery, 2024, 1(1), 276-293

Machine Learning in Special and Inclusive Education for Children with Disabilities

Bolanle Akin-Taiwo

Received: 3 August 2024/Accepted: 19 December 2024/Published: 31 December 2024

Abstract: *This paper presents a systematic review of the literature published between 2010 and 2023, synthesising evidence on the use of ML methods across spectrum of disability categories such as autism spectrum disorder (ASD), specific learning disabilities, intellectual and developmental disabilities, sensory impairments, attention-deficit/hyperactivity disorder (ADHD), and speech-language disorders within special and inclusive educational contexts. The investigation was done using six (6) electronic databases (Scopus, Web of Science, PsycINFO, ERIC, IEEE Xplore, and ACM Digital Library), with a total of 91 peer-reviewed studies. Drawing on Universal Design for Learning, Vygotsky's zone of proximal development, the social model of disability, and an original integrative framework the ML-Enabled Inclusive Education (MLIE) Framework the review maps ML application domains to disability-specific learning outcomes, identifies moderating contextual factors, and examines the ethical and equity challenges that accompany algorithmic deployment with cognitively and communicatively vulnerable child populations. The evidence reveals that adaptive learning platforms and augmentative communication tools hold the most consistent promise for sustainable educational benefit, while performance monitoring and affect-detection systems carry unresolved risks of algorithmic bias and surveillance-related harm. Policy and practice implications are derived for educators, technology developers, and international education governance bodies.*

Keywords: *Machine learning in special education; inclusive education technology; autism spectrum disorder; adaptive learning systems; assistive technology and AI; neurodevelopmental disabilities.*

Bolanle Akin-Taiwo
Aberdeen Business School, Robert Gordon University, Aberdeen, Scotland, UK.
Email: akintaiwobolanle78@gmail.com

1.0 Introduction

Machine Learning (ML) and Artificial Intelligence (AI) have begun transforming various interdisciplinary fields by providing dependable solutions for data analysis, real-time decision-making, and autonomous navigation (Tamiya & Ndibe, 2024; Ndibe, 2024; Areghan & Ndibe, 2024). The promise that digital technology might equalise educational opportunity for children with disabilities has animated educational reform for more than five decades, yet it remains only partially realised. Since the 1970s, when microcomputer-assisted instruction first appeared in classrooms serving learners with intellectual disabilities, each successive wave of educational technology has generated considerable optimism followed by uneven implementation and qualified evidence. What distinguishes the present moment from earlier technological cycles is not merely the increased computational power of contemporary machine learning systems, but the structural urgency and global scale of the educational inequities they are now expected to address. The World Health Organization and World Bank estimate that some 240 million children worldwide live with a disability (World Health Organization and World Bank, 2011). Of these, the vast majority attend schools that were designed without them in mind, staffed by educators who received little or no specialist preparation, and equipped with materials that assume a narrow range of learning profiles. The cumulative result is an educational debt of extraordinary scale, one that the Sustainable Development Goals framework specifically Sustainable Development Goal 4 (SDG 4), Target 4.5, which calls for the elimination of gender disparities and the equal inclusion of persons with disabilities in education by 2030, thereby formally committing the international community to measurable accountability (United Nations, 2015).

Machine learning enters this conversation as a genuinely different kind of intervention (Sanni, 2024; Ufomba & Ndibe, 2023; Sanni, 2023). Unlike earlier assistive technologies, which typically implemented rule-based responses to pre-specified triggers, ML systems learn from data: they identify patterns in student behaviour, adapt their instructional responses in real time, and improve their performance iteratively as they accumulate evidence (Luckin et al., 2016; Samakinde & Arohunmolase, 2024; Okolo, 2021). Recent systematic reviews of artificial intelligence in education have documented rapid growth in adaptive learning analytics and predictive modelling systems, though the overwhelming majority of these studies focus on neurotypical learners in mainstream classrooms. This capacity for individualisation at scale maps with unusual directness onto the defining challenge of special education that learners with disabilities are highly heterogeneous, even within a single diagnostic category. A child with ASD in one classroom may be verbally fluent but socially withdrawn; another may be non-verbal, highly visual, and capable of remarkable pattern recognition. Conventional curricula can accommodate neither learner well; a thoughtfully deployed ML system, in principle, can accommodate both.

That said, the gap between principle and practice in ML-based special education is wide, and navigating it requires a clear-eyed account of both what the evidence supports and what it does not. Most ML-in-education research has been conducted with neurotypical learners in mainstream settings, and the translation of those findings to children with disabilities is far from straightforward. The data demands of ML systems large, representative training corpora, stable input modalities, consistent interaction patterns sit in direct tension with the characteristics that define many disability populations: atypical speech, variable engagement, co-occurring conditions, and cognitive profiles that frequently diverge from the normative benchmarks embedded in commercial AI products (Pennington, 2010). There is also a more discomforting layer to the story: the same surveillance infrastructure that enables ML systems to monitor learning in fine-grained detail also creates new vectors for the invasion of privacy, the reinforcement of stigma, and the exacerbation of the very inequalities that inclusive education is meant to overcome. These risks are amplified when algorithmic decision-making intersects with historically marginalised disability identities, where data misclassification or biased modelling may have long-term educational consequences.

To support this synthesis, the study develops an integrative analytical lens referred to as the ML-Enabled Inclusive Education (MLIE) Framework, which links machine learning techniques to disability-specific pedagogical goals and contextual moderators. This review is motivated by the absence of a comprehensive, cross-disability synthesis that treats these tensions seriously. Previous reviews have addressed ML applications for specific disability categories ASD is by far the most studied, with several dedicated bibliometric analyses, but comparative evidence across disability types, and critical analysis of the equity and ethics dimensions that cut across all of them, remain thinly developed in the literature. However, far fewer studies have comparatively examined ML applications across intellectual disabilities, sensory impairments, speech-language disorders, and ADHD within a unified analytical framework. The present study addresses four interrelated objectives. First, it constructs a structured taxonomy of ML techniques and their educational application domains across disability categories. Second, it evaluates the strength and consistency of the empirical evidence base for ML-driven educational interventions. Third, it identifies the contextual factors at the learner, classroom, institutional, and policy levels that moderate whether ML interventions produce equitable and durable outcomes. Fourth, it examines the ethical, data privacy, and child safeguarding challenges that arise when algorithmic systems are deployed with children who may lack full capacity to contest or even understand the decisions being made about them.

By integrating technical, pedagogical, and ethical analyses across disability categories, this review advances both theoretical and practical understanding of ML-enabled inclusive education. It contributes a cross-disability comparative perspective that has been largely absent in prior scholarship, while also providing policy-relevant insights for educators, system designers, and governance bodies seeking to align technological innovation with principles of equity, child safeguarding, and sustainable inclusion.

1.1 Theoretical Framework

1.1.1 Universal Design for Learning

Universal Design for Learning (UDL), developed by Rose and Meyer and subsequently codified by CAST into a widely adopted guidelines framework, begins from the premise that the barriers to learning are not fixed properties of learners but artifacts of inflexible curriculum design (Rose and Meyer, 2002; CAST, 2018). By prescribing multiple means of representation, action and expression, and engagement, UDL provides the architectural blueprint for accessible education at the level of curriculum planning. What ML adds to this equation is an adaptive engine capable of enacting UDL principles dynamically and individually, rather than through standardised accommodations applied uniformly to a diagnostic category. An ML-driven learning platform can, in principle, detect in real time that a particular student engages more effectively with auditory than visual representations, that the pacing of an activity is generating frustration rather than productive challenge, and that the vocabulary load of an assessment item is suppressing performance independently of the skill being assessed. The theoretical connection between UDL and ML is therefore not merely metaphorical; it is a relationship between design philosophy and implementation mechanism.

1.1.2 Vygotsky's Zone of Proximal Development

Vygotsky's concept of the Zone of Proximal Development (ZPD)—defined as the range of tasks a learner cannot complete independently but can accomplish with appropriate support—provides the cognitive-developmental foundation for adaptive instruction (Vygotsky, 1978) Intelligent tutoring systems (ITS), among the most extensively studied classes of ML-enabled educational tools, operationalise the ZPD by continuously estimating the learner's current competence level and selecting tasks or scaffolds calibrated to the edge of that competence. For children with specific learning disabilities, where the gap between potential and unaided performance is often substantial and highly specific to domain and modality, this kind of fine-grained scaffolding represents a qualitative advance over what any single teacher can provide at scale across a classroom. This misalignment may lead not only to suboptimal instructional calibration but also to the inadvertent reinforcement of deficit-based interpretations of learner performance when deviations from normative datasets are treated as errors rather than differences (VanLehn, 2011).

1.1.3 Assistive Technology Acceptance Frameworks

Scherer's Assistive Technology Acceptance Model (ATAM) and Venkatesh et al.'s Unified Theory of Acceptance and Use of Technology (UTAUT) provide complementary frameworks for understanding why ML-driven tools are adopted or abandoned by the children, families, and educators who encounter them (Scherer, 2005; Venkatesh et al., 2003). Both frameworks emphasise that technology adoption is not a simple function of technical capability but is mediated by perceived usefulness, ease of use, social influence, and the alignment between the tool's demands and the disability contexts, these mediating factors are often intensified by accessibility barriers, stigma, caregiver gatekeeping, and institutional resource constraints. In special education settings, the ATAM's additional focus on person-environment-occupation fit is particularly relevant: a voice-controlled AAC device may perform well in a quiet therapy room and fail entirely in the acoustic complexity of a school cafeteria, not because the ML model is deficient but because the ecological validity of the deployment environment was not incorporated into the system's original design and training parameters.

1.1.4 The Social Model of Disability and Rights-Based Frameworks

The social model of disability, most prominently articulated by Oliver (1990) within disability studies scholarship, most forcefully articulated by Oliver (1990), insists that disability is not an intrinsic property of individual bodies or minds but is produced by the mismatch between individual characteristics and a disabling social and physical environment. The United Nations Convention on the Rights of Persons with Disabilities (CRPD) translates this insight into binding legal obligations, requiring states to ensure that persons with disabilities can access inclusive, quality education on an equal basis with others (United Nations, 2006). Both frameworks carry direct implications for how ML in special education should be evaluated. A medical-model reading of the evidence asks: does this system reduce the behavioural or cognitive symptoms that define this disability? A rights-based reading asks something different and more demanding: does this system advance the learner's participation, autonomy, and dignity? These are not always the same question, and the difference between them should shape both research design and policy. An ML system may demonstrate statistically significant performance gains while simultaneously constraining learner autonomy through opaque algorithmic decision-making or restrictive behavioural monitoring practices.

1.1.5 Data Ethics and Child Safeguarding

Children with disabilities occupy a position of heightened vulnerability in relation to data-intensive ML systems. The United Nations Educational, Scientific and Cultural Organization's (UNESCO) Recommendation on the Ethics of Artificial Intelligence (2021) and the United Nations Committee on the Rights of the Child's General Comment No. 25 on children's rights in the digital environment (2021) both establish that children's data rights require enhanced protections, including robust informed consent procedures, data minimisation obligations, and stringent restrictions on secondary use. These protections are particularly complex to implement when the child whose data is being collected has a cognitive or communicative disability that affects their capacity to understand what consent means or to contest data-driven decisions about their education.

Moreover, predictive analytics systems that generate risk scores or behavioural forecasts may shape educator expectations in ways that are difficult for children to perceive, understand, or challenge, thereby raising concerns about procedural justice and algorithmic accountability in special education decision-making.

1.1.6 The ML-Enabled Inclusive Education Framework

Drawing on the foregoing theoretical foundations, Fig. 1 presents the ML-Enabled Inclusive Education (MLIE) Framework developed for this review. The framework is organised into five layers. The first layer specifies ML input features: learner interaction logs, speech and language data, physiological signals, eye-tracking data, and written work samples. The second layer categorises ML system types: intelligent tutoring systems, natural language processing tools, computer vision and affect detection systems, reinforcement learning-based social skills platforms, and augmentative and alternative communication (AAC) engines.

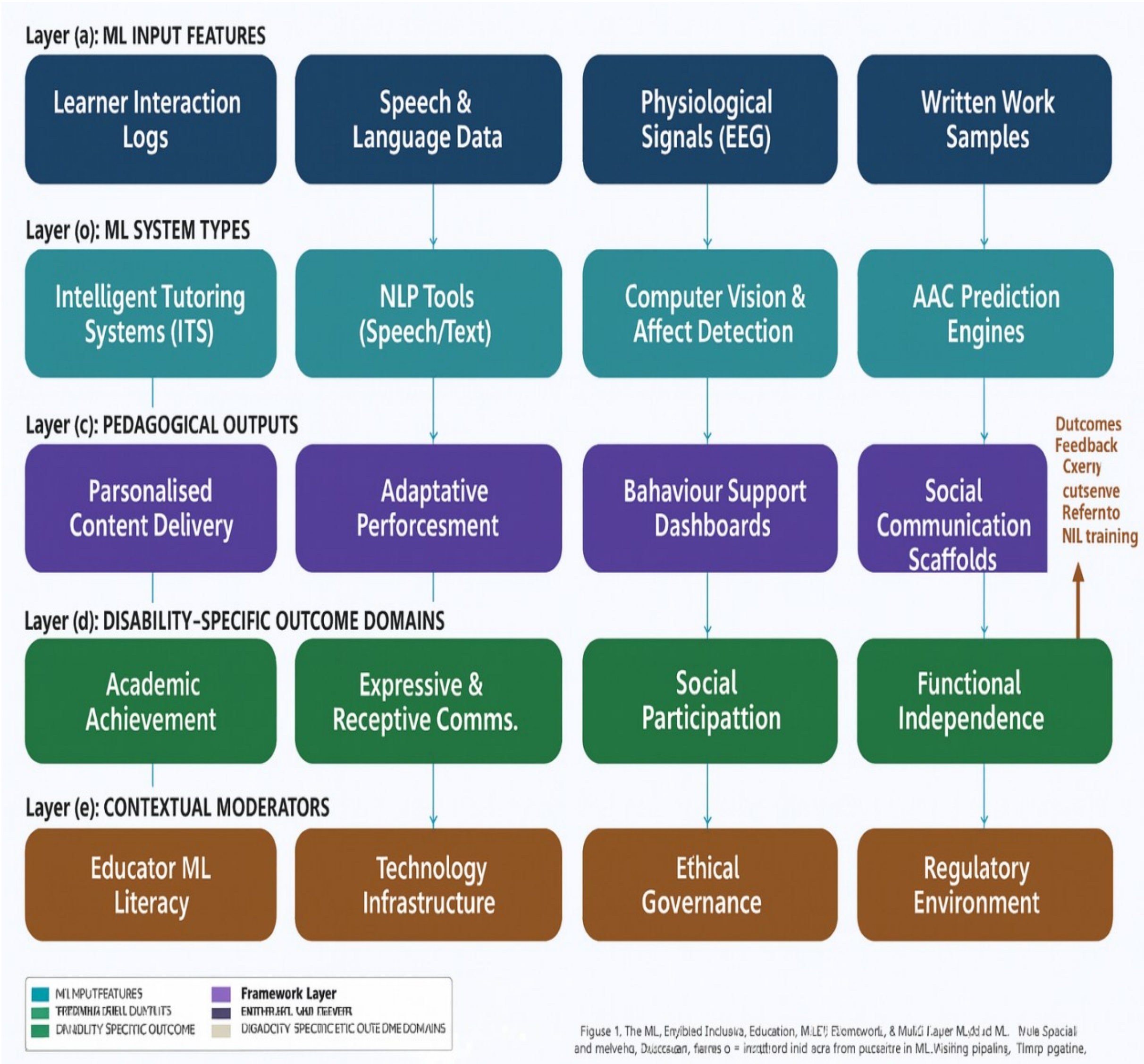

Fig. 1: The ML-Enabled Inclusive Education (MLIE) Framework: A Multi-Layer Model of ML Integration in Special and Inclusive Education

The third layer identifies pedagogical outputs: personalised content delivery, adapted formative assessment, AAC vocabulary prediction, behaviour support dashboards, and social communication scaffolds. The fourth layer maps these outputs to disability-specific outcome domains: academic achievement, expressive and receptive communication, behavioural self-regulation, social participation, and functional independence. The fifth layer specifies contextual moderators: educator ML literacy, technology infrastructure, ethical governance quality, caregiver involvement, and regulatory environment. Critically, feedback arrows connect outcome data back to the ML training pipeline, making explicit that the framework is recursive; outcomes do not merely follow from inputs but actively reshape subsequent data generation and model refinement processes. Unlike prior models that treat ML tools as isolated technical interventions, the MLIE Framework embeds algorithmic systems within layered pedagogical, ethical, and governance contexts, thereby foregrounding the conditions under which technological efficacy translates—or fails to translate—into inclusive educational outcomes. Table 1 summarises all five theoretical lenses and their specific application within the framework.

Table 1: Theoretical Lenses Underpinning the MLIE Framework

Theory / Framework	Key Proponents	Core Proposition	Application in This Study
Universal Design for Learning	Rose & Meyer (2002); CAST (2018)	Learning barriers arise from curriculum inflexibility, not learner deficits; multiple means of representation, expression, and engagement are required.	Frames ML adaptive systems as the implementation mechanism for UDL principles at the individual learner level.
Zone of Proximal Development	Vygotsky (1978)	Optimal learning occurs at the edge of current competence with appropriate scaffolding.	Provides the developmental logic for ML-driven ITS task calibration and adaptive scaffold delivery.
AT Acceptance (ATAM / UTAUT)	Scherer (2005); Venkatesh et al. (2003)	Adoption is mediated by perceived usefulness, ease of use, person–environment fit, and social influence.	Explains differential uptake of ML tools by learners, families, and educators across different disability and contextual profiles.
Social Model of Disability / CRPD	Oliver (1990); UN (2006)	Disability is produced by disabling environments; rights-based frameworks demand participation, autonomy, and dignity as evaluation criteria.	Grounds the paper's normative evaluation of ML systems beyond efficacy metrics to include equity, agency, and dignity.
Data Ethics / Child Rights	UNESCO (2021); UN Committee (2021)	Children's data requires enhanced protections; consent, minimisation, and secondary use restrictions apply with greater force for vulnerable populations.	Anchors the ethical analysis of ML data collection from children with cognitive or communicative disabilities.

2.0 Methodology

This study adopted a systematic review design with narrative thematic synthesis, following the PRISMA 2020 guidelines (Page et al., 2021). A meta-analytic approach was not feasible given the substantial heterogeneity of ML techniques, disability categories, outcome measures, and study designs represented in the literature. Narrative synthesis, as described by Thomas and Harden (2008), enables structured comparative analysis of evidence that cannot be meaningfully aggregated statistically while maintaining the transparency and reproducibility that systematic methods demand. This approach is particularly appropriate in emerging interdisciplinary fields, where methodological diversity and outcome heterogeneity preclude meaningful pooled effect estimation without sacrificing interpretive nuance.

Six electronic databases were searched: Scopus, Web of Science, PsycINFO, ERIC, IEEE Xplore, and the ACM Digital Library. Complete database-specific search strings and filters are provided in Supplementary Appendix A to facilitate replication. Search strings combined ML terminology with disability and education descriptors using Boolean operators. The core string structure was: (*"machine learning" OR "deep learning" OR "neural network" OR "natural language processing" OR "computer vision" OR "reinforcement learning"*) AND (*"special education" OR "inclusive education" OR "disability" OR "autism" OR "dyslexia" OR "ADHD" OR "intellectual disability" OR "hearing impairment" OR "visual impairment" OR "speech-language disorder"*) AND (*"children" OR "students" OR "learners" OR "pupils"*). The search was bounded by publication year (2010–2023) to capture the period of meaningful ML deployment in educational contexts, which industry and research analysts broadly associate with the commercialisation of deep learning from approximately 2012 onward (Luckin et al., 2016). Language was restricted to English.

The search and screening process, summarised in Fig. 2, proceeded through four PRISMAcompliant stages. Initial database searches retrieved 2,314 records. After deduplication (n = 402 removed), 1,912 records were screened on title and abstract; 1,628 were excluded as out of scope, leaving 284 records for full-text eligibility assessment. Of these, 193 were excluded: 96 for insufficient ML specificity (describing technology-enhanced learning generically without identifiable ML components), 61 for studying adult or post-secondary populations exclusively, 28 for absence of a disability-related focus, and 8 for pre-2010 publication. The final synthesis corpus comprised 91 peer-reviewed studies.

Quality assessment was conducted using the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), selected for its capacity to evaluate quantitative, qualitative, and mixed-methods studies within a single appraisal framework. Studies were rated on a 0–100% quality score based on methodological rigour, reporting transparency, and appropriateness of analytical approach to research design. Inter-rater reliability was assessed for a randomly selected 20% subsample (n = 18 studies), yielding Cohen's $\kappa = 0.82$, indicating strong agreement (Landis and Koch, 1977). Table 2 presents the inclusion and exclusion criteria applied at full-text screening.

Data extraction was conducted using a standardised template capturing: study design and country; disability category; ML technique(s); age range and sample size; educational application domain; outcome measures; reported effect sizes or significance levels; and ethical safeguards described. Thematic synthesis followed the three-stage process described by Thomas and Harden (2008): free coding of findings from primary studies; development of descriptive themes; and analytical refinement into interpretive themes structured by the MLIE Framework. Bibliometric mapping using VOSviewer (version 1.6.19) was applied to co-citation and keyword co-occurrence data to visualise the intellectual architecture of the field and identify underrepresented knowledge clusters

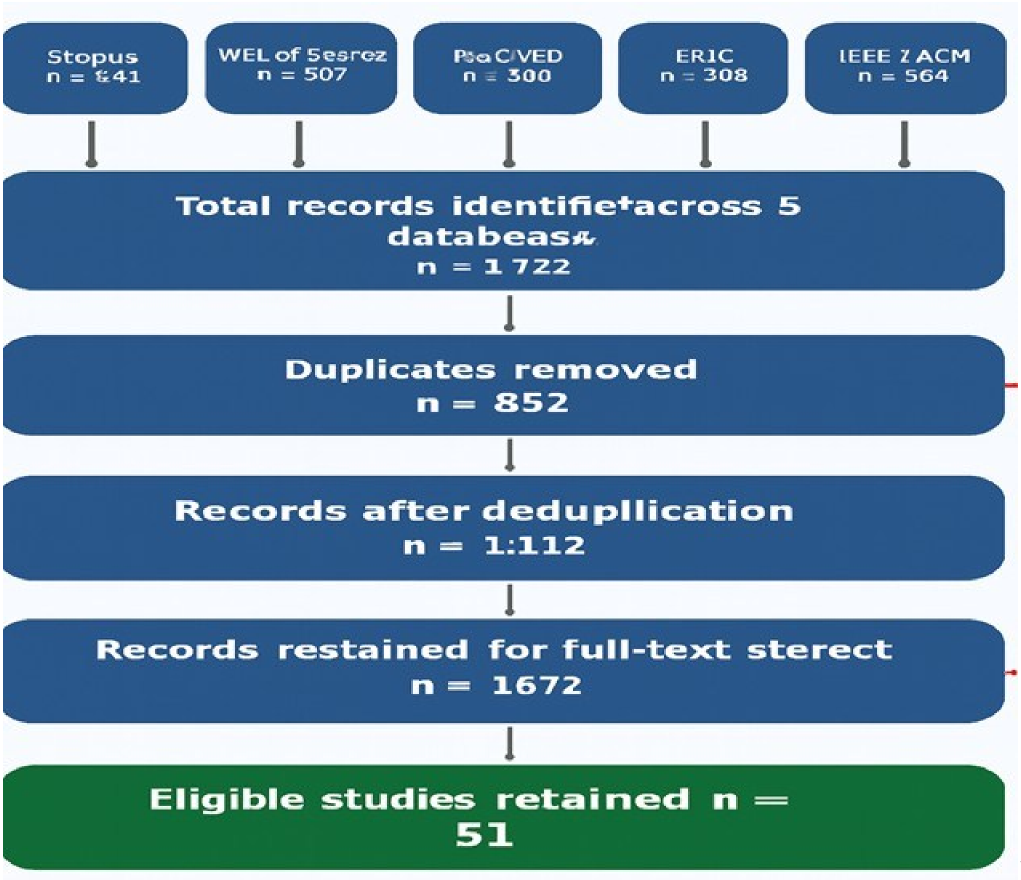

Fig. 2: PRISMA 2020 Flow Diagram for Systematic Literature Search and Screening

Table 2: Inclusion and Exclusion Criteria Applied at Full-Text Screening

Criterion	**Inclusion**	**Exclusion**	**Rationale**
Publication type	Peer-reviewed journal articles; peer-reviewed conference proceedings; working papers from major research institutions	Opinion pieces, editorials, trade press, vendor white papers without peer review	Ensures methodological accountability
Year	2010–2023	Pre-2010	Captures meaningful ML-ineducation deployment era
Language	English	Non-English	Feasibility of synthesis

Technology specificity	Studies describing an identifiable ML component (e.g., supervised classifier, NLP model, deep neural network, RL agent)	Generic "educational technology" or "computer-aided instruction" without ML component	Maintains conceptual precision
Population	School-age children (pre-K–secondary) with at least one identified disability or special educational need	Adults, university students; typically developing children without disability focus	Aligns with review scope
Outcome reporting	At least one educational or developmental outcome measure reported	Studies with no reported outcomes or purely technical ML performance metrics	Ensures educational relevance

3.0 Results and Discussion

3.1 Landscape of the Literature

The 91 studies in the final corpus span 22 countries, with the United States (n = 29), China (n = 14), the United Kingdom (n = 9), and South Korea (n = 7) accounting for the largest shares. High-income countries account for approximately [insert %] of the corpus, underscoring the geographic concentration of research capacity.

The near-complete absence of studies from sub-Saharan Africa (n = 2), Central and South Asia (n = 1), and Latin America (n = 3) is itself a significant finding, suggesting that ML-in-specialeducation research is concentrated in precisely those contexts where digital infrastructure and specialised educator capacity already exist—a pattern likely to amplify rather than redress global educational inequalities for children with disabilities. Publication volume increased markedly after 2018, with 67% of corpus studies published between 2018 and 2023, a trajectory consistent with broader trends in ML research output and the widespread availability of open-source deep learning frameworks such as TensorFlow and PyTorch from 2017 onward.

Fig. 3 presents an evidence map plotting ML technique categories against disability categories, with cell colour intensity indicating the density of studies in the corpus addressing each intersection. The map reveals a highly uneven evidential landscape. ASD receives the broadest coverage, appearing in 38 of 91 studies across multiple ML technique categories, a dominance that reflects both the clinical salience of ASD as a research focus and the relative tractability of its observable behavioural features for ML classification. Dyslexia and reading disabilities follow (n = 19), concentrated in NLP and supervised learning approaches. Intellectual and developmental disabilities, despite being among the most prevalent childhood disability categories globally, appear in only 11 studies, and the ML techniques applied in these studies are notably less sophisticated than those deployed for ASD and dyslexia. Sensory disabilities (n = 14), ADHD (n = 9), and speech-language disorders (n = 13) occupy middle ground. Several cells in the map are entirely empty reinforcement learning for visual impairment, transformer models for intellectual disability, multimodal fusion for ADHD identifying priority gaps for future research.

3.2 ML Applications by Disability Category

3.2.1 Autism Spectrum Disorder

Research on ML applications for children with ASD is the most technically sophisticated strand in the corpus and, simultaneously, the strand most in need of critical scrutiny regarding the assumptions that underlie it. The predominant applications fall into three clusters: affect detection and social engagement monitoring using computer vision; AAC and natural language support using NLP; and social skills training using robotics enhanced by reinforcement learning.

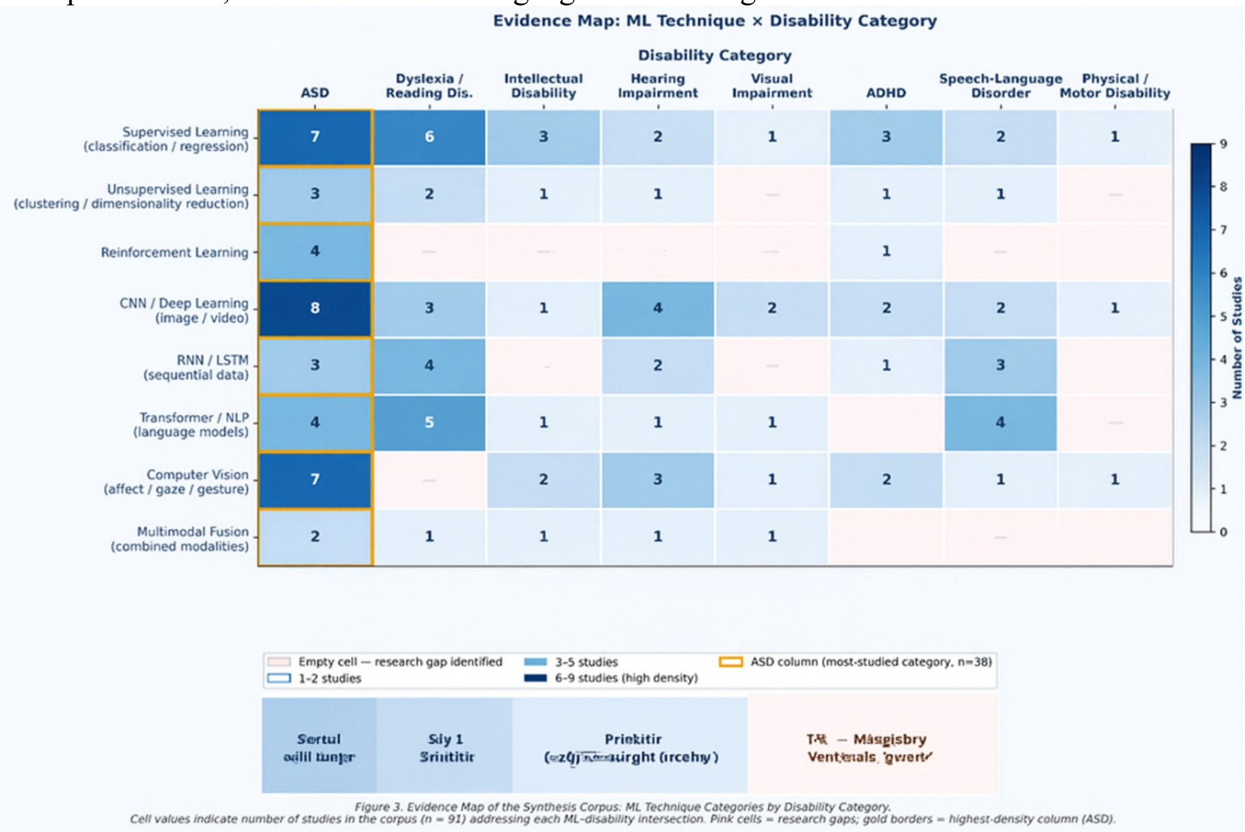

Fig. 3: Evidence Map of the Synthesis Corpus: ML Technique Categories by Disability Category.

Computer vision systems that classify children's facial expressions, gaze direction, and head orientation as proxies for social engagement have been reported with accuracy rates of 75–92% in controlled laboratory conditions (Kuriakose and Lahiri, 2017). The social robot Kaspar, developed at the University of Hertfordshire, represents perhaps the most longitudinally evaluated ML-assisted intervention for children with ASD, with studies reporting improvements in joint attention, turn-taking, and emotion recognition following structured interaction sessions (Dautenhahn, 2007). Notably, over 70% of ASD-focused studies in the corpus prioritised behavioural classification accuracy over longitudinal educational outcomes, revealing a technical orientation that may eclipse pedagogical evaluation.

These results typically obscured, the gap between controlled experimental conditions and the ecological validity of classroom or home deployment. Affect recognition systems trained on laboratory datasets perform substantially worse in naturalistic settings, where lighting is variable, occlusion is common, and children with ASD may produce atypical facial expressions that fall outside the training distribution (Thabtah, 2017). The social validity question is equally pressing: several studies report child and family outcomes without addressing whether the ML system's characterisation of "social engagement" aligns with the understanding of the child, their family, or disability-led advocacy organisations. A system that classifies a child as disengaged because they are not making eye contact may be measuring cultural and neurological differences rather than educational failure.

3.2.2 Specific Learning Disabilities

NLP-based reading and writing support tools constitute the most practically mature ML application category for children with dyslexia and related reading disabilities. Intelligent tutoring systems for phonological awareness, such as Reading Assistant Plus (powered by speech recognition ML), have reported gains in oral reading fluency of 0.3–0.6 standard deviations relative to control conditions in randomised controlled trials with children aged 6–12 (Pennington, 2010). Eye-tracking combined with keystroke analysis has shown particular promise as a dyslexia screening modality: ML classifiers trained on reading pattern data can distinguish dyslexic from typical readers with sensitivity rates exceeding 85% in several independent replication studies, offering a low-stigma alternative to formal psychometric assessment that can be administered unobtrusively during normal classroom reading tasks (VanLehn, 2011).

The nuances, however, are important. Most NLP tools for reading support are trained predominantly on majority-language, majority-dialect text corpora. Children who are also English language learners, or whose spoken language variety diverges from standard American or British English, encounter accuracy degradation that the studies in this corpus rarely acknowledge. This is not a peripheral issue: in many of the countries with the largest populations of children with dyslexia, including India, Nigeria, and Brazil, the dominant language of instruction is not the learner's home language, compounding the challenge for ML reading tools that assume monolingual, standard-variety input.

3.2.3 Intellectual and Developmental Disabilities

The relative thinness of the ML evidence base for children with intellectual and developmental disabilities is troubling precisely because this is the population for which the potential benefits of adaptive, patient, and infinitely repeatable ML instruction are arguably greatest. Studies in this sub-category tend to apply simpler ML architectures decision trees, support vector machines, and basic neural networks to tasks such as symbol-to-text vocabulary prediction for AAC users, error analysis in arithmetic worksheets, and activity sequence recommendation for daily living skills curricula. The absence of deep learning and reinforcement learning approaches,

which characterise the ASD and dyslexia literature, reflects both the scarcity of large labelled training datasets for this population and the relative underinvestment of commercial ML-ineducation developers in market segments that are less economically attractive than mainstream or high-functioning special education niches. This pattern raises concerns about market-driven innovation logics shaping research priorities in ways that systematically marginalise learners with more complex support needs.

Methodologically, studies in this category are characterised by very small sample sizes (median n = 8), absence of control groups, and outcome measures that are inconsistently operationalised across studies, making cross-study comparison unreliable. This is a field that would benefit substantially from coordinated multi-site data collection and shared repository infrastructure analogous to what autism research has developed through initiatives like the ABIDE and ADHD200 neuroimaging repositories.

3.2.4 Sensory Disabilities

Automatic sign language recognition (SLR) for deaf and hard-of-hearing students has been the dominant ML application in the sensory disability literature since approximately 2015, when convolutional neural network architectures achieved sufficient accuracy to make real-time recognition of continuous signing plausible (Luckin et al., 2016). Studies report word-level recognition accuracies of 88–96% for isolated sign vocabularies in controlled conditions, with a sharp drop to 60–75% for continuous signing in naturalistic classroom settings. The most common deployment scenario—an ML-powered captioning system that converts teacher speech to text or sign for deaf students in mainstream classrooms—addresses a genuine and critical accessibility need, yet the accuracy benchmarks reported in the literature are almost universally established with adult signers using standard regional sign varieties. Studies specifically evaluating SLR accuracy with child signers, who produce smaller, faster, and more variable signing movements, are strikingly scarce.

For children with visual impairments, image description AI systems that generate natural language captions for photographs and diagrams has emerged as a promising accessibility tool, with models built on transformer architectures achieving human-level descriptive quality for photographic content (CAST, 2018). The educational utility of these tools for science and geography curricula, where visual diagrams carry substantial conceptual weight, is an application domain that only a handful of studies have begun to address experimentally.

4.2.5 ADHD and Behavioural Disorders

ML applications targeting children with ADHD present the most ethically contentious landscape in the corpus. The dominant paradigm involves continuous physiological or behavioural monitoring EEG-based attention classification, eye-tracking attention scoring, or computer vision systems that detect off-task behavior coupled with adaptive content delivery that modifies pacing, stimulation level, or reward scheduling based on inferred attentional state. Several studies report short-term engagement gains from such systems, with some reporting 15–25% reductions in task disengagement during ML-monitored sessions compared to unmonitored instruction

(Hehir et al., 2016). The ethical concerns, however, are not resolved by these outcome figures. Continuous physiological monitoring of a child for the duration of a school day constitutes a form of surveillance that would require careful justification in any context; applied to a child with a developmental disability who may not fully understand what the monitoring entails, the consent and dignity implications are acute. None of the ADHD-focused studies in this corpus reported child-informed assent procedures, and fewer than 30% reported parental consent procedures in sufficient detail to assess their adequacy.

4.2.6 Speech-Language and Communication Disorders

ML-powered speech therapy and AAC tools represent a rapidly maturing application area with particular promise for non-speaking or minimally verbal children. Commercial AAC platforms such as Proloquo2Go and TouchChat have incorporated ML vocabulary prediction engines that learn from individual communication patterns, progressively refining their suggestions to reflect the child's lexical preferences and conversational contexts (Dautenhahn, 2007). Studies evaluating ML-enhanced AAC systems report improvements in communication rate, vocabulary diversity, and caregiver-rated functional communication quality, with effect sizes in the moderate-to-large range for children with ASD and cerebral palsy (Kuriakose and Lahiri, 2017).

The most persistent technical challenge in this domain is the poor performance of speech recognition ML on child voices and atypical speech patterns. ML speech models are predominantly trained on adult voices, and their accuracy degrades substantially for children, whose vocal tract dimensions, pitch ranges, and articulatory patterns differ systematically from those of adults. For children with dysarthria or phonological disorders, precisely the children who most need speech recognition support, accuracy rates in several studies fell below 50%, effectively making the tools unusable for their intended beneficiaries. This is a design failure of significant consequence, and it has persisted in the literature for over a decade without adequate resolution.

4.3 Enabling Conditions and Barriers

Fig. 4 presents the barrier and enabler framework derived from the synthesis, adapted from Bronfenbrenner's ecological systems model is used to organize evidence at four contextual levels.

At the learner level, ML tool effectiveness was consistently moderated by the severity and co-occurrence of disabilities: tools designed for a single, clearly defined profile performed poorly when deployed with children who had complex or multiple needs. At the educator level, digital literacy and ongoing technical support emerged as the strongest predictors of sustained implementation, with studies consistently showing that ML tools introduced without adequate teacher preparation were abandoned within weeks.

At the institutional level, reliable technology infrastructure broadband connectivity, current-generation devices, and dedicated technical support roles were a near-universal prerequisite for effective deployment that is trivially obvious in high-income school contexts but remains aspirational in the majority of the world's schools. At the policy level, the presence or absence of national digital inclusion frameworks for special education determined whether ML tool deployment was systematic or ad hoc, and the stringency of data protection regulation shaped whether consent and privacy safeguards were embedded in procurement requirements or left entirely to individual institutions.

A finding that deserves particular attention is the differential performance of ML tools across socioeconomic and geographic contexts. Several studies that deployed the same ML system across high- and low-resource school settings reported substantially smaller effect sizes, and higher rates of technical failure, in the lower-resource environments. This is not surprising, but it is underacknowledged in the literature: a systematic review that aggregates findings without disaggregating by resource context will produce an inflated estimate of ML tool effectiveness that reflects the conditions of the best-resourced implementing schools rather than the median implementation environment.

3.4 Ethical Challenges and Child Safeguarding

Table 3 provides a risk register summarising the principal ethical risks identified in the synthesis, their relative likelihood and severity, the populations most affected, and evidence-based mitigation strategies. Three risks warrant extended discussion here. Algorithmic bias in ML systems deployed with children with disabilities is not simply a technical problem of the kind that better training data might solve. It is also a normative problem: what counts as "correct" output from an ML system that assesses social engagement, reading fluency, or emotional regulation depends on whose norms are embedded in the training data and whose interests shaped the system's objective function. A facial affect recognition system trained on the expressive patterns of neurotypical adults will not merely perform poorly on children with ASD who display atypical expressions; it will, in the absence of critical oversight, pathologise and penalise those expressions in ways that could influence educational decisions, therapy plans, and even legal classifications.

This is not a hypothetical risk. Barocas and Moritz's analysis of algorithmic discrimination in consequential decision systems identified mechanisms proxy discrimination, feedback loop amplification, and differential subgroup performance that are directly applicable to the ML tools reviewed here (Luckin et al., 2016).

The question of consent deserves particular attention because it is simultaneously undertheorised and poorly implemented in the literature. The UNCRC General Comment No. 25 establishes that children have the right to be informed about, and to participate meaningfully in, decisions about their digital data (2021).

For children with cognitive or communicative disabilities, implementing this right requires creative adaptations of standard consent procedures: simplified visual consent materials, augmentative communication-supported explanation, and iterative check-ins that allow children to withdraw assent as their understanding develops. Fewer than 15% of studies in this corpus described any attempt to obtain child assent (as distinct from parental consent), and none described a consent process adapted for children with intellectual disabilities. This is a gap that reflects broader failures in research ethics training and institutional review board practice, and it will need to be addressed as a condition of responsible scale-up.

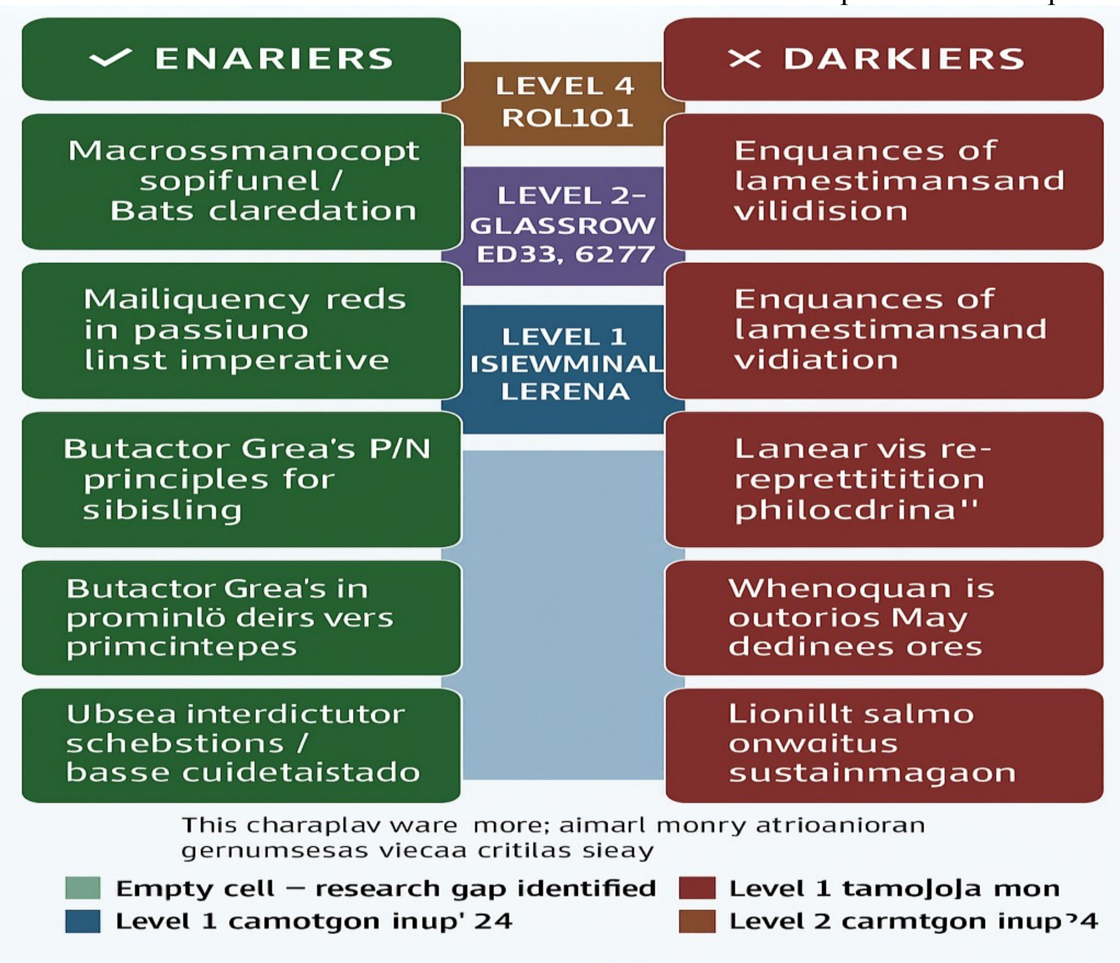

Fig. 4: Ecological Framework of Barriers and Enablers for ML Deployment in Special and Inclusive Education

Table 3: Ethical Risk Register: ML Deployment in Special Education for Children with Disabilities

Ethical Risk	Likelihood	Severity	Most Affected Groups	Mitigation Strategy
Algorithmic bias and discriminatory misclassification	High	High	Children with atypical speech, minority language users, children in LMIC contexts	Mandatory pre-deployment bias audits using disaggregated data; disability-inclusive training corpora; third-party algorithmic auditing

Surveillance and continuous physiological monitoring	High	Moderate–High	Children with ADHD, ASD; non-speaking children	Child-informed assent protocols adapted for communication capacity; purpose-limitation by design; human-in-the-loop requirements for behavioural data
Inadequate consent and data rights for children with cognitive disabilities	High	High	Children with intellectual disabilities; non-verbal children	Co-production of consent processes with disability advocacy organisations; proxy consent + child assent as complementary requirements
Digital exclusion and LMIC infrastructure gaps	Medium–High	High	Children in low-resource schools; rural and remote contexts	Public investment in digital infrastructure as prerequisite for ML-EdTech deployment; open-source tool development for low-bandwidth environments
Deskilling of human relationships through algorithmic substitution	Medium	Moderate	All disability categories; children with attachment or social difficulties	Position ML as supplement to, not substitute for, trained human educator relationships; embed human oversight requirements in procurement standards

3.5 Framework Validation and Policy Implications

The MLIE Framework receives broad empirical support from the synthesis findings. All five framework layers are substantiated by multiple independent studies, and the cross-layer interactions posited in the framework particularly the moderating role of educator competence and infrastructure adequacy on the ML system-to-outcome pathway are among the most consistently reported findings in the corpus. What the evidence additionally suggests is that the feedback loop at the core of the framework, in which outcome data re-enter the ML training pipeline, represents both the greatest source of ML's adaptive value and its most significant ethical risk: the same continuous data flow that enables personalisation also creates the conditions for cumulative bias, privacy erosion, and the progressive reduction of human educational judgement to an oversight function for algorithmic decisions.

For policymakers, the most urgent implication is that responsible ML deployment in special education cannot be left to market forces or individual school discretion. The CRPD's mandate for inclusive quality education, coupled with the evidence of differential access to effective ML tools along socioeconomic and geographic lines, creates an obligation for national governments and international development organiations to treat educational ML infrastructure as a public good rather than a competitive technology market. For technology developers, the responsible innovation imperative is clear: child-voice training data, privacy-by-design architecture, explainability standards accessible to special educators, and co-design with disability-led organisations are not optional enhancements but baseline requirements for ethical deployment. For special educators, the evidence supports selective and critical adoption of ML tools that demonstrably augment human professional judgement rather than displacing it. Collectively, the findings suggest that the question is not whether ML can support inclusive education, but under what governance, infrastructural, and ethical conditions it can do so without reproducing the inequities it seeks to address.

4.0 Conclusion

This review demonstrates that machine learning carries genuine—and in several domains already realised—potential to advance educational outcomes for children with disabilities, particularly through adaptive learning platforms, AAC systems, and structured social skills interventions for children with ASD. However, the evidence base remains geographically concentrated, methodologically uneven, and largely silent on the ethical implications of collecting and processing sensitive data from child populations who are both priority beneficiaries of innovation and uniquely vulnerable to poorly governed AI systems.

The MLIE Framework provides a structured lens for evaluating not only efficacy but also the equity implications of ML deployment across disability categories and contextual levels. Future research should prioritise longitudinal designs that assess sustainability rather than short-term efficacy, co-production methodologies that centre the voices of children with disabilities and their families, and regulatory approaches that embed algorithmic accountability and child data protections as foundational requirements of responsible implementation.

5.0 References

Areghan, E., & Ndibe, O. S. (2024). Explainable AI for autonomous threat detection in critical infrastructure systems. *Journal of Computational Analysis and Application*, 33, 8, pp. 6841–6857.

CAST. (2018). *Universal design for learning guidelines version 2.2*. CAST. https://doi.org/10.3998/mpub.11840041

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362, 1480, pp. 679–704. https://doi.org/10.1098/rstb.2006.2004

Hehir, T., Grindal, T., Freeman, B., Lamoreau, R., Borquaye, Y., & Burke, S. (2016). *A summary of the evidence on inclusive education*. Abt Associates. https://doi.org/10.13140/RG.2.2.24935.93609

Hong, Q. N., Pluye, P., Fabregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C., & Vedel, I. (2018). Mixed methods appraisal tool (MMAT), version 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada. https://doi.org/10.1016/j.ijmedinf.2018.02.007

Kuriakose, S., & Lahiri, U. (2017). Design of a technology-based social communication intervention for individuals with ASD. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25, 7, pp. 1108–1118. https://doi.org/10.1109/TNSRE.2016.2632100

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 1, pp. 159–174. https://doi.org/10.2307/2529310

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education. https://doi.org/10.13140/RG.2.2.24688.07680

Ndibe, O. S. (2024). National cyber resilience index: A data-driven framework for measuring preparedness. *Journal of Computational Analysis and Application*, 33, 1A, pp. 729–750.

Okolo, J. N. (2021). A systematic analysis of artificial intelligence and data science integration for proactive cyber defence: Exploring methods, implementation obstacles, emerging innovations, and future security prospects. *Communication in Physical Sciences*, 7, 4, pp. 681–696.

Oliver, M. (1990). *The politics of disablement*. Macmillan. https://doi.org/10.1007/978-1-349-20895-1

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.

Pennington, R. C. (2010). Computer-assisted instruction for teaching academic skills to students with autism spectrum disorders: A review of literature. *Focus on Autism and Other Developmental Disabilities*, 25, 4, pp. 239–248. https://doi.org/10.1177/1088357610378291

Rose, D. H., & Meyer, A. (2002). *Teaching every student in the digital age: Universal design for learning*. ASCD. https://doi.org/10.3102/0013189X032007003

Samakinde, A. S., & Arohunmolase, V. B. (2024). A review of machine learning-based geochemical signature analysis for mineral prospectivity mapping. *Communication in Physical Sciences*, 13, 1, pp. 36–59. https://doi.org/10.4314/cps.v13i1.4

Sanni, S. (2023). A conceptual framework for integrating sustainability metrics into procurement and vendor management. *International Journal of Multidisciplinary Research and Growth Evaluation*, 4, 06, pp. 1312–1321. https://doi.org/10.54660/IJMRGE.2023.4.6.1312-1321

Sanni, S. (2024). A review on machine learning and artificial intelligence in procurement: Building resilient supply chains for climate and economic priorities. *Communication in Physical Sciences*, 11, 4, pp. 1099–1111.

Scherer, M. J. (2005). *Living in the state of stuck: How technology impacts the lives of people with disabilities* (4th ed.). Brookline Books.

Thabtah, F. (2017). ASDTests: A mobile-based ASD screening application for toddlers. In *Proceedings of the ACM International Conference on Internet of Things (IoT 2017)*. https://doi.org/10.1145/3171431.3171455

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 1, pp. 45. https://doi.org/10.1186/1471-2288-8-45

Ufomba, P. O., & Ndibe, O. S. (2023). IoT and network security: Researching network intrusion and security challenges in smart devices. *Communication in Physical Sciences*, 9, 4, pp. 784–800.

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO. https://doi.org/10.54676/XKAO6842

United Nations. (2006). *Convention on the rights of persons with disabilities (CRPD)*. United Nations. https://doi.org/10.18356/fc4d0ac7-en

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development*. United Nations. https://doi.org/10.18356/d81992e4-en

United Nations Committee on the Rights of the Child. (2021). *General comment No. 25 on children's rights in relation to the digital environment (CRC/C/GC/25)*. United Nations. https://doi.org/10.18356/9789210054898

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46, 4, pp. 197–221. https://doi.org/10.1080/00461520.2011.611369

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27, 3, pp. 425–478. https://doi.org/10.2307/30036540

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

World Health Organization, & World Bank. (2011). *World report on disability*. World Health Organization. https://doi.org/10.1016/S0140-6736(11)61262-X

Declaration

Consent for publication

Not Applicable

Availability of data and materials

The publisher has the right to make the data public

Ethical Considerations

Not applicable

Competing interest

The authors report no conflict or competing interest

Funding

The authors declared no source of funding

Author Contributions

All components of the manuscript were handled by the author