International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 255-273
https://doi.org/10.70917/ijcisim-2026-2517
Article

Designing an Explainable Multimodal AI Framework for Identifying Creative Artistic Developmental Potential in Early Childhood: A Theoretical and Methodological Foundation

Shyamala B.[1], Sridhara P. Acharya[2]

[1] Institute of Computer Science and Information Science, Srinivas University, Mangaluru, Karnataka, India.
Email: shyamala.sanil89@gmail.com
ORCID: 0009-0006-0288-063X
[2] Institute of Computer Science and Information Science, Srinivas University, Mangaluru, Karnataka, India.
Email: shyamala.b@srinivasuniversity.edu.in
ORCID: 0000-0002-9996-0014

Abstract: Discovering artistic-creative potential in children in an early stage has remained critical even so largely underexplored research domain in educational research. Existing computational approaches in talent recognition primarily focus on academic advancement or developmental delays and disorders. This paper presents the theoretical and methodological foundation for the Explainable Multimodal Artificial Intelligence for Talent Identification (EMAI-T) framework. EMAI – T is an interdisciplinary effort designed to enhance transparent and theory -driven recognition of artistic talent in children at their early stage. The proposed study initially outlines the design and preparation of a multimodal reference dataset from experts in the art field capturing textual, visual, behavioural of creativity across different age groups. Later, the study implements a conceptual and theoretical framework by integrating Gardener's Multiple Intelligence Theory, Gagne's Differentiated Model of Giftedness and talent (DMGT), and Explainable AI(XAI) tenets to translate psychological dimensions into suitable computational interpretations. Ultimately, the paper introduces the architectural design of the EMAI-T framework, which adopts explainability techniques such as SHAP, LIME, and attention mechanism to ensure transparency and trust in AI interpretations.

Keywords: Multimodal learning, Explainable AI(XAI), Artistic creativity, Early childhood education, Talent Recognition, Human-centred AI, Educational Technology

1. Introduction

1.1 Overview and Significance

Children in their early stage represents a crucial period for the fostering of imagination, creativity, and cognitive elasticity. These are the essential skills that strengthens foundation for sustained learning, innovation, and emotional well-being [1][2]. Despite the fast arising of artificial intelligence (AI) in education, most of the applications focus primarily on growth related screening and the identification of developmental delays, learning difficulties rather than the recognizing emotional, creative -behaviour, leadership and socio- emotional potential in children [3][4]. Though teacher-based observations in traditional education system play a valuable role, however remain restricted by subjectivity, limiting the fair identification of diverse forms of skills and talent. Consequently, there is a time-sensitive need to re-envision the role of AI not as an only diagnostic tool for deficits, but as a facilitator for identifying, analysing and fostering creative skills, artistic expression, and comprehensive development in young learners aged 8-12.

Existing research mainly applies advancement of AI to identify either learning disabilities, developmental delays, autistic behaviour or predict academic performances using standardized datasets that neglect creative and socio-emotional behaviours [3][4]. Recent review of multimodal learning analytics in education furthermore highlights that, in spite of rapid methodological advances, major applications that prioritize identifying creative potential in children, interpretability, and context of early development remain limited ([13]. Studies that integrate multimodal learning analytics with creativity assessment also remain rare and insufficient. Drawing on Gardner's (2011) theory of multiple intelligence and Kaufman and Sternberg's (2023) framework of creative potential, the proposed study examines how well an AI- driven observation can identify early indicators of creative and artistic articulations children aged 8-12.

1.2 Research Gap

A systematic review of recent studies uncovered four significant research gaps that restrict the advancement of AI-based artistic-creative talent identification in early childhood contexts

(1) Lack of Conceptual and Systematic Foundation

AI – based assessment models often prioritize developmental delays, disorder detection rather than the anticipatory, strength-based classification. There is also absence of theoretical grounding in developmental psychology in children, and few have studies attempted to interpret psychological constructs such as creativity, giftedness, or artistic potential into computationally interpretable representations. The absence of a theoretical bridge between psychological constructs and AI modelling limits the scientific and ethical validity of current systems.[5][6].

(2) Lack of Diverse, Multimodal and Longitudinal Data

Available AI studies extensively rely upon a short-term, single-modality datasets such as textual or performance based. The single modality and short-term datasets often fail to capture the diversity complexity and abundance of creative, expressive, and emotional behaviour of children. Recent extensive reviews of multimodal learning analytics suggest that, although there is as major advances in data fusion techniques, yet most studies remain bound by limited modality diversity, short observation windows, and controlled experimental settings, with insufficient attention to early childhood developmental contexts [13]. The lack of demographically diverse and longitudinal multimodal datasets also limits fairness, generalizability, interpretability and applicability in AI- based creativity and talent supporting systems [7][8].

(3) Limited Interpretability, Trust and Transparency in AI Models

Most of the AI models in early education or child psychology frequently function as black boxes. The black box AI models generate results and prediction without clear and valid reasons. The transparency level in these black box AI models undermines trust in AI driven recommendations among parents and educators especially in subjective domains such as creativity and artistic potential. Recent comprehensive reviews of multimodal learning analytics in education identify limited transparency and interpretability as sustained barrier to the integration of AI systems in educational settings, particularly in contexts involving children.[13]. Hence there is a strong need for Explainable AI(XAI) techniques such as SHAP, LIME, and attention mechanism that ensure interpretability, transparency and human-understandable insights [9][10].

(4) Limited Ethical, Technological, and Structural Readiness for Real-World implementation

There is lack of ethical and practical concerns including privacy, informed consent, data bias, and infrastructural limitations that hinder real world implementation and validation. The lack of technological preparedness, expert evaluation, teacher training, and governance frameworks in educational institutions requires responsible deployment of AI systems in creative learning and talent development contexts [11-12].

Addressing these major research gaps requires an interdisciplinary approach that combines developmental psychology, educational technology, and explainable AI to encourage responsible, transparent, and creativity-centred talent identification in early childhood.

1.3 Problem Statement

The proposed is study is highly motivated in the basis that every child exhibits a unique and multi-layered talents that span beyond standard academic metrics and creative behaviour in children deserve equal opportunity in recognition and nurturing. Research in creativity highlights that human capability encompasses originality, effectiveness, and various forms of expression that includes artistic behaviour, socio-emotional, and distinct thinking capacities [14].

In contrast, many modern AI- driven educational systems fortify a deficit-oriented approach by focusing on the disorder and lacks in children rather than focusing on what ability or strength the child is having. Many existing AI applications in education predominantly prioritize academic performance prediction, STEM capability, drop-out risk identification, learning delays identification and mitigation of learning gaps [15]. Such approaches may often frame young learners primarily in terms of measurable limitations, thereby regenerating deficit-focused models long criticized in educational research [16]. As a result, non- academic strengths and creative capabilities that withstand standardized quantification risk being marginalised.

While AI systems are increasingly competent of processing complex multimodal data, existing systematic reviews indicate sustained limitations, including data set bias, limited representativeness, fairness concerns, and insufficient model interpretability [15][17]. In addition, the ethical and transparent frameworks necessary for strength-oriented AI deployment in education remain underexplored. Consequently, early talent identification in children remains disintegrated and predominantly academically centred.

These limitations reinforce the need for a theoretically grounded, multimodal, and explainable AI framework that is capable of supporting strength-based and ethically responsible early talent identification.

This study is motivated by the vision that AI must function not solely as a predictive or corrective mechanism, but as a strength-oriented, psychologically informed system that supports comprehensive development in children. By integrating established creativity theory into AI design [14] and incorporating explainable modelling approaches [17], this research aims to bridge the gap between technological innovation and human -centered educational development.

1.4 Research Objectives and Scope

This paper intends to initiate a theory grounded and methodologically well-structured foundation for an Explainable Multimodal AI framework for early childhood artistic talent identification. Precisely, it proposes an interdisciplinary conceptual model, a well curated multimodal data set design blueprint and an explainable AI architectural usability.

Research Objectives

The paper addresses the following objectives

Objective 1: To design a multidisciplinary theoretical foundation for identification creative artistic developmental potential

The proposed objective develops a conceptually aligned framework that integrates developmental psychology, creativity research, giftedness theory, and explainable artificial intelligence (XAI). The proposed framework intended to translates of artistic-creative potential constructs into computationally interpretable representations by drawing on Gardner’s Multiple Intelligence Theory and Gagne’s Differentiated Model of Giftedness and talent (DMGT) making it more suitable for multimodal AI modelling.

The objective ensures, theoretical coherence, construct validity, and alignment between psychological meaning and computational representation.

Objective 2: To formulate a theory-aligned and ethically governed blueprint of the multimodal dataset

The proposed objective articulates the structured, methodological requirement to construct a multimodal dataset that captures artistic-creative expression in children aged 8-12 years. The design framework indicates the selection of theoretically grounded modalities such as textual responses, visual, behavioural data, expert- informed description, construct alignment methods and ethical protocols which include informed consent and data governance mechanism.

Rather than implementing quantitative data collection, the objective defines a scalable blueprint capable of supporting model training, future generalizability studies.

Objective 3: To outline the architectural and explainability principles of EMAI-T framework

The proposed objective aligns a theory-grounded multimodal AI architecture capable of integrating multiple expressive data streams while maintaining interpretability. The framework specifies modality -specific representation strategies and cross-modal fusion design, alignment between psychological constructs and computational layers, explainability mechanism such as SHAP, LIME, attention-based modelling and transparency requirements for stakeholder interpretability.

This objective mainly addresses the need for fairness, accountability trust in AI- based talent identification system.

Scope of the Study

The proposed study is limited to conceptualization and methodological design. It does not include dataset construction, model training, quantitative validation and pilot deployment. These phases establish subsequent empirical studies.

While the proposed framework is adaptable across multiple domains of talent, this paper focuses particularly on artistic-creative potential in children aged 8-12 years as an initial implementation scope.

1.5 Research Contribution

The paper primary contributes to the fields of artificial intelligence in education, creativity, research, and explainable AI

1. Theoretical Contribution

By integrating developmental psychology and giftedness theory with explainable AI principles, the paper advances a strength-based computational interpretations of artistic-creative potential. This bridges a major gap between psychological theory and AI system design.

2. Methodological Contribution

By embedding fairness, diversity, and ethical governance at the conceptual stage, the paper proposes a structured multimodal dataset design framework addressing persistent challenges related to bias and representativeness in AI- driven educational systems.

3. Architectural Contribution

The paper outlines a transparent and interpretable multimodal XAI architecture with the techniques SHAP, LIME, attention mechanism and ensures computational outputs remain understandable to educators and parents.

4. Ethical and Design Contribution

The paper shifts AI based deficit assessment of a child from deficit-oriented profiling towards strength-based assessment, positioning explainability and psychological validity as foundational design requirements.

Together these contributions establish a theoretically consistent and ethically grounded blueprint for responsible AI integration in early childhood talent identification.

1.6 Overview of the Proposed Framework

The Explainable Multimodal AI for talent Identification (EMAI-T) framework is envisioned as a theory-based, explainability -driven architecture for the early developmentally responsive analysis of early of artistic-creative potential in children aged 8-12 years. The framework combines psychological theory, multimodal data representation, and interpretable artificial intelligence within an integrated conceptual structure.

A theory-grounded construct operationalization layer lies at its core, informed by developmental psychology and well-established models of intelligence and giftedness, including Gardner's Multiple Intelligence Theory (Gardner, 1983, 1999) and Gagne's Differentiated Model of Giftedness and Talent (Gagne, 2015). These theoretical perspectives play a major role in guiding the identification of artistic -creative indicators and enable their translation into computationally interpretable constructs. The EMAI- T framework embeds construct validity directly into system design, ensuring alignment between psychological meaning with computational representation.

The proposed framework further integrates a multimodal data layer representing complementary dimensions of various creative expression. These include visual artifacts such as drawing, gesture dynamics, and affective -behavioural cues such as hand movement, posture, pauses, drawing patterns. The framework reflects contemporary advances in multimodal learning by integrating heterogeneous expressive channels that demonstrate enhanced representational richness through cross-modal integration. [18]. Each modality is encoded through specialized representation mechanisms in prior to integration.

The reference-based modelling orientation is the defining feature of the EMAI-T framework. Domain -expert artistic reference data are conceptually applied to establish anchor representations within a shared multimodal creativity embedding space. Early developmental child expressions are mapped into this space to enable similarity-based alignment. This approach highlights structural correspondence rather than stylistic replication, supporting developmentally sensitive interpretation rather than explicit labelling.

Within an explainable modelling layer, multimodal inference is conducted that integrates modality specific representations through structure fusion while preserving construct-level traceability. Attention based relational mechanisms and post-hoc interpretability techniques enable transparent feature attribution and ensure that the outputs remain analytically interpretable and educationally meaningful by ensuing trust. Explainability techniques such as SHAP [20] and LIME [21] empower feature- level attribution and interpretable reasoning. This ensures that model outputs are not limited predictive, but analytically transparent and pedagogically meaningful.

Figure 1, illustrates the EMAI-T framework aligns psychological theory, multimodal data representation, reference-based embedding, and explainable AI within a systematic architectural logic.

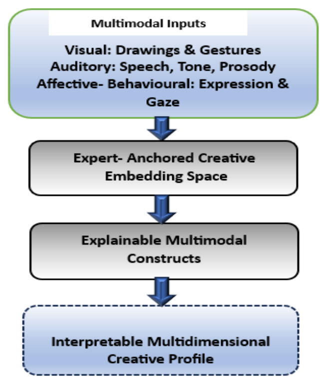

Figure 1: Multimodal pipeline for early artistic-creative profiling [source: self]

The present study states the theoretical and methodological foundation of the proposed framework, while data -driven dataset construction and validation constitute subsequent phases of the broader research program.

Section 3 enhances this conceptual overview by specifying the methodological design, construct operationalization processes, and multimodal modelling architecture in detail.

2. Conceptual and Theoretical Foundation

The conceptual foundation of the Explainable Multimodal AI for Talent Identification (EMAI-T) framework is established in a multidisciplinary integration of, education, developmental psychology, giftedness theory, and explainable artificial intelligence (XAI). Aligned with Objective 1 of this study, this section develops a theoretically consistent and construct-valid basis for translating artistic-creative potential into computationally interpretable representations making suitable for multimodal AI modelling. The proposed framework embeds established psychological models into system design to ensure alignment between developmental meaning and computational representation [22][14]

2.1 Integrating Developmental and Educational Psychology with AI

Research studies in developmental and educational psychology formulate creativity and talent as multidimensional and context-sensitive [23-24]. Artistic creative potential in childhood extends beyond academic performances and includes expressive, affective dimensions of intelligence [1][2]. Applications of AI in educational contexts have historically highlighted performance prediction and cognitive outcomes, often overlooking creative and socio-emotional indicators [25]. The EMAI-T framework adopts a strength-based perspective, to ensure theoretical coherence and construct validity and conceptualize early creative behaviours as indicators of latent [24]. Within this approach, multimodal AI modelling is guided by psychologically grounded definitions of artistic- creative development.

2.2 Gardner's Multiple Intelligence Theory and Artistic –Creative Indicators

Multiple Intelligence (MI) theory by Howard Gardner reconceptualizes intelligence as a multidimensional construct integrating discreate domains, including spatial, bodily-kinaesthetic, musical, interpersonal, and intrapersonal intelligences [1]. Spatial, bodily- kinaesthetic, and intrapersonal intelligences are particularly prominent in early artistic talent recognition as they reflect visual imagination, expressive movement, and emotional awareness.

Mapping between theoretical constructs and observable modalities enabled by the domain-specific clarity of MI theory. Such as visual-spatial intelligence may be reflected in compositional structure and symbolic representation in drawings, bodily-kinaesthetic intelligence in gesture dynamics and movement patterns and intrapersonal intelligence in affective modulation and reflective engagement [1].

By translating these modality-linked indicators into computational features, the EMAI-T frameworks ensures that Multimodal data acquisition and representation are theoretically anchored rather than purely data-driven.

2.3 Gagne's Differentiated Model of Giftedness and Talent (DMGT)

Gagne's Differentiated Model of Giftedness and Talent (DMGT) distinguish between natural abilities(giftedness) and systematically developed competencies(talent), emphasizing the role of intrapersonal and environmental catalysts in transforming potential into performance (Gagne 2004,2025) This developmental distinction provides a structured framework for interpreting early creative behaviours.

Within EMAI-T, giftedness is conceptualized as latent creative potential observable through spontaneous artistic behaviours, while talent represents the progressive development of these abilities through sustained engagement and contextual support. This distinction informs a hierarchical modelling logic in which early multimodal indicators correspond to latent potential and integrated behavioural patterns reflect developmental progression [24].

2.4 Explainable Artificial Intelligence as a Translational Mechanism

Although deep learning architectures offer powerful pattern recognition capabilities, they often lack interpretability and trust [26] [27]. While explainable artificial Intelligence (XAI) provides the methodological bridge between psychological constructs and computational inference. Transparency and accountability are critical in educational contexts particularly those involving children [25].

The proposed EMAI-T framework incorporates explainable AI techniques such as LIME [26], SHAP [27] and attention -based interpretability mechanism [28]. The framework aims to attribute model predictions to specific multimodal features. These explainable techniques enable stakeholders to interpret and understand how visual, expressive, visual, and affective indicators contribute to talent related inferences.

The framework aligns with emerging principles of trustworthy and ethical AI in education by integrating XAI as a foundational design component [29].

2.5 Integrative Conceptual Model

Integration of ML theory, DMGT, and XAI principles forms a three-dimensional conceptual model that links multidimensional definitions of intelligence, developmental trajectories of talent formation, and transparent computational inference.

MI theory defines various the domains of artistic-creative intelligence (Gardner, 1983), DMGT frames these abilities within a developmental progression from giftedness to talent (Gagne, 2015), and XAI ensures that multimodal modelling processes remain interpretable and accountable [27]. These components together establish a strength-based, human-centered framework in which psychological validity and computational transparency are mutually reinforcing.

The following section builds upon this theoretical foundation by outlining the methodological principles guiding multimodal dataset design.

3. Methodological Design and Multimodal Dataset Blueprint

Consistent with Objective 2, this section outlines the methodological design principles guiding the development of a theoretically grounded, ethically governed multimodal reference dataset for artistic- creative talent identification.

Rather than implementing empirical data collection at this stage, the present work defines the structural, theoretical, and governance framework required for constructing a robust dataset capable of supporting explainable multimodal AI modelling. The section further describes the conceptual dataset construction pipeline including participant selection strategies, multimodal data capture modalities, and expert annotation protocols which together inform the future integration and modelling processes of the proposed framework.

3.1 Dataset Design Philosophy and Study Rationale

The methodological foundation of the EMAI-T framework is premised on the recognition that artistic -creative talent manifests through heterogeneous, multimodal expressions encompassing visual, auditory, behavioural, and affective dimensions [14]. Traditional educational datasets often relay on unidimensional metrics such as standardized test scores or linguistic outputs, which inadequately capture expressive and embodied forms of intelligence [24]

To address this limitation, the proposed dataset blueprint adopts a multimodal learning paradigm, in which complementary sensory and behavioural streams are integrated to model complex human expression [78]. The approach corresponds with contemporary advances in multimodal machine learning that demonstrate improved representational richness and contextual sensitivity through cross -modal integration.

The blueprint of the dataset is conceptually designed to serve multiple purposes

1) A training dataset collected from domain experts for multimodal XAI modelling

2) A benchmarking resource dataset for future research in early childhood education.

The structure of the dataset blueprint is adopted by Gardner's Multiple Intelligences Theory (Gardner, 1983, 1999) and Gagne's DMGT (Gagne, 2004, 2015) by ensuring theoretical alignment between psychological constructs and computational representation.

The proposed study adopts a theory-grounded dataset design strategy to capture multimodal indicators of artistic creativity. The dataset design pipeline supports the construction of both a multimodal reference dataset collected from domain experts and a child dataset used during pilot implementation with participants aged 8-12. The figure 2 illustrates methodological workflow guiding this dataset construction process and illustrates participant cohorts, creative task design, multimodal data capture and dataset organization.

Figure 2. Dataset Construction Pipeline [source: self]

3.2 Multimodal Data Modalities and Representational Blueprint

The dataset blueprint consists of three primary modality clusters that are textual, visual, and behavioural. Each modality corresponds to theoretically grounded indicators of artistic-creative expression.

(a) Textual Modality

The text -based modality is captured by participants responses to a structured questionnaire which provide self-reported information which are related to artistic talent, creative thinking including artistic experience, creativity interest, motivation and training in prior. This modality is grounded in creativity theory and psychometric assessment, which completes the visual and behavioural modalities by capturing underlying characteristics that are not observable directly from the final drawing image itself.

Textual modalities represented at three hierarchical levels

1)Response level which contains the encoded answers to individual questionnaire sets.

2)Attribute level where related responses are aggregated into artistic talent constructs

3) Talent representation level which provided a comprehensive indicators of the participants artistic aptitude for multimodal fusion.

(b) Visual Modality

The visual modality captures artifact-based and process-based indicators of artistic-creative expression through structured drawing, sketching, painting, and digital design tasks. In alignment with visual –spatial intelligence (Gardner, 1983, 1999) and the developmental emphasis of DMGT (Gagne, 2004,2015), the dataset blueprint conceptualizes visual creativity as a multi-layered construct encompassing perceptual diversity, compositional organization, and generative exploration.

Visual representation is structure across three complementary levels:

1) Low- level perception features

These captures basic ow-level descriptors such as colour, intensity, edges, texture, gradients, corners, stroke thickness and compositional balance. These describes the raw appearance of the final drawing

2**) Mid- level compositional complexity indicators**

These captures mid -order metrics such as shapes and structural patterns such as contours, geometric shapes, object parts, symmetry, spatial dispersion measures, composition and drawing layout by combining low level features.

3**)High-level visual features**

These levels capture semantic and conceptual information including object identity, artistic style, aesthetic quality, structural asymmetry, proportional variation. These are abstract representations learned by deep learning models or derived from expert interpretations

[c] Behavioural Modality

The behavioural modality captures expressed, emotional, and regulatory correlates of creative engagement through drawing process video-based multimodal analysis. The behavioural modality grounded in affective computing and multimodal learning analytics research. This modality conceptualizes creative potential not merely as final output, but as an emergent state characterized by emotional variability, expressive embodiment, and sustained engagement.

Representation is structured across three subdimensions:

1) Process- Based Dynamics

In the process-based dynamics drawing processes are digitally captured. Stroke velocity variation, pressure variability, revision frequency, and spatial exploration trajectories are incorporated to model generative experimentation. In early childhood talent identification process-level modelling is important where creative potential often manifests through exploratory behaviour rather than final artifact sophistication.

2) Embodied Movement and Postural Expressivity

Embodied movement and postural expressivity capture gesture amplitude, movement fluidity, posture shifts, lean-in behaviour, and kinematic variability which serve as proxies for embodied immersion and expressive confidence.

3) Engagement and Self- Regulation Signals

Time -on-task persistence, distraction frequency, task re-engagement latency, and observable self-regulatory gestures are incorporated as behavioural indicators of intrinsic motivation and creative flow states.

Table 1: Multimodal dataset Design Blueprint

Modality	Data Type	Theoretical Basis	Example Features	Processing Strategy	Ethical Consideration

Textual	Questionnaire responses	Psychometric Assessment Theory	Creative disposition, self-reported artistic experience	NLP BERT	Anonymized responses
Visual	Final Drawings, Sketches	MI -Spatial intelligence	Drawing behaviour, Stroke dynamics, symmetry, colour variability	CNN- based feature extraction	Anonymized artifacts
Behavioural	Drawing process video	DMGT- Intrapersonal catalysts	Engagement, gaze, gesture fluidity	Pose estimation/ affect modelling	Anonymized clips
Metadata	Age, background	DMGT- Environmental catalysts	Demographic distribution	Fairness auditing	GDPR compliance

3.2.1 Multimodal Fusion Architecture

The proposed dataset blueprint adopts a temporally synchronized multimodal fusion framework to support integrative modelling while preserving theoretical traceability. Each modality is first represented through modality-specific encoding layers that generate structured embeddings. Temporal synchronization windows enable the alignment of concurrent visual, auditory, and affective-behavioural signals, facilitating relational modelling of cross-modal expressive dynamics.

Rather than relying exclusively on early -stage feature concatenation, the framework conceptually supports hybrid fusion strategies combining intermediate joint-embedding alignment with late-stage construct-level aggregation. This layered integration approach enables both representational richness and interpretive transparency, as construct-level indicators remain traceable to modality-specific contribution.

Attention-based cross-modal mechanism may further allow the system to identify dominant expressive channels during creative engagement episodes, thereby supporting context-sensitive analysis of multimodal creativity patterns. Importantly, the architecture prioritizes transparency and interpretability over opaque optimization, consistent with the framework’s commitments to explainable AI in educational decision-support contexts.

To conceptually illustrate how modality-specific representation are integrated, the proposed multimodal fusion architecture is presented in Figure 3.

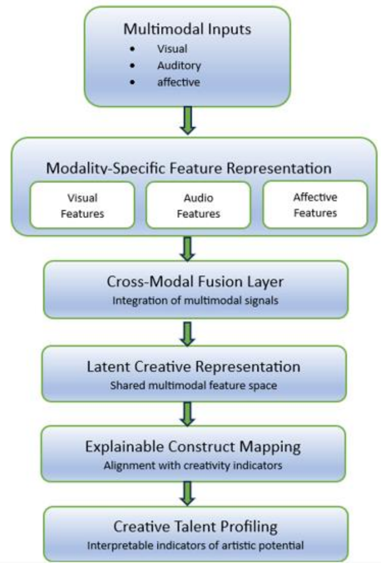

Figure3: Conceptual multimodal fusion architecture illustrating the integration of textual visual, behavioural signals for interpretable artistic talent profiling [source: self]

The multimodal fusion architecture conceptually integrates modality-specific feature representations derived from visual, auditory, and affective signals. These representations are combined within a shared latent space to enable interpretable modelling of creative indicators relevant to early artistic talent identification

3.2.2 Cross-Modal Construction Alignment

Creativity constructs relevant to early artistic expression may manifest through multiple observable

Table 2: Cross-Modal Alignment of Creativity Constructs and Multimodal Indicators

Creativity Construct	Visual Indicators (Drawing/ Gesture)	Textual Indicators (Question answer responses)	Behavioural Indicators
Divergent Thinking	Novel composition, originality compositional asymmetry, object diversity	Self-reported creative interests, curiosity Imagination	Fluency, exploration of alternatives, revisions, Flexibility
Expressive Fluency	Stroke velocity variation, colour diversity, richness and completeness	Self-reported ease, confidence in expression, idea through art	Continuous stroke, smooth drawing flow, speed, number of expressed elements, Gesture amplitude, movement fluidity
Imaginative Symbolism	Symbolic imagery, metaphorical representations, Structural recombination patterns	Sel reported imagination, fancy, symbolic thinking, preferences for imaginative activities	Evolution of symbolic concepts during drawing, transformation of ideas
Intrinsic Motivation	Revision persistence, exploratory trajectory	Self-reported Preferences for creating art without expectation,	Time-on-task, lean-in behaviour

		persistence, willingness to draw during free time, enjoyment during art activity	
Emotional Responsiveness	Dynamic colour shifts, emotional tone, expressive use of visual elements	Self-reported emotional sensitivity, emotional engagement with artistic activity	Change in drawing rhythm, persistence, behavioural signs of engagement

Note: The table presents a conceptual alignment between creativity constructs and observable multimodal indicators. The indicators represent theoretical guidance for dataset design rather than finalized computational features. A more granular operational specification of features will be developed in the subsequent dataset development study.

3.2.3 Reference-Based Multimodal Representation and Modelling Paradigm

Beyond defining modality-specific representations, the EMAI-T framework is conceptually oriented toward a reference-based multimodal modelling paradigm. Rather than treating artistic talent identification as a categorical prediction problem, the proposed system is designed to construct a structured creativity embedding space anchored by domain-expert expressive signatures.

Within this paradigm, multimodal data derived from expert artists serve as anchor representations of crystallized artistic competencies across visual, auditory, and affective – behavioural dimensions. These expert-derived embeddings establish a multidimensional reference manifold reflecting theoretically grounded constructs of divergent thinking, expressive fluency, symbolic imagination, and intrinsic creative engagement.

Multimodal expressive profiles of children are subsequently projected into the same construct- aligned embedding space, enabling similarity based relational analysis. Importantly, similarity is conceptualized at the level of structural creative indicators such as variability, expressive modulation, and cross-modal coherence rather than stylistic imitation or aesthetic conformity. This distinction preserves developmental sensitivity while mitigating normative bias.

3.3 Annotation Framework and Theoretical Validation

The proposed annotation framework is grounded in operationalized dimensions derived from Multiple Intelligence theory and the Differentiated Model of Giftedness and Talent (DMGT). Expert raters including art educators and creativity researchers are conceptually positioned as evaluators of theoretically defined indicators such as novelty, elaboration, flexibility, expressive modulation, compositional balance, and sustained creative engagement [14]. These indicators are aligned with broader constructs of divergent thinking, expressive fluency, symbolic imagination, and intrinsic motivation as outlined in the cross-modal construct alignment framework.

Annotation strategies incorporate structured rating scales (e.g., Likert-type measures) alongside categorical and qualitative tagging to capture both quantitative gradations and contextual nuances of creative expressions. The design blueprint includes provisions for interrater reliability assessment (e.g., Cohen's Kappa, Intraclass Correlation Coefficients) to ensure annotation consistency and construct stability in subsequent empirical phases.

3.4 Data Preprocessing and Multimodal Integration Preparation

The data preprocessing framework is designed in a systematic structure to ensure data quality, construct preservation, and cross-modal compatibility within the proposed multimodal AI architecture. Standardized data preprocessing procedures includes data cleaning, normalization and feature scaling are incorporated at the design level to reduce noise, and distributional inconsistencies.

Feature representation strategies are aligned with modality-specific encoding paradigms. Visual artifacts processed through convolutional architectures to capture spatial structural patterns. Textual modalities are processed through Natural Language Processing techniques such as BERT and behavioural signals may be modelled using pose-estimation and behavioural analysis frameworks to encode embodied engagement. Preprocessing is conceptualized not merely as technical conditioning, but as a construct-preserving transformation stage that safeguards alignment between psychological meaning and computational representation.

The proposed design facilitates flexible fusion strategies for cross- modal integration such as early-stage feature integration, intermediate joint embedding alignment, and late- stage construct level aggregation consistent with contemporary multimodal learning paradigm. As a result, integrated representation forms a structured multimodal embedding space capable of modelling interdependencies across expressive, affective, and behavioural dimensions.

3.5 Ethical Governance and Data Protection Framework

Ethical governance constitutes a foundational component of the EMAI-T dataset blueprint. As mentioned, the involvement of minors and the collection of multimodal behavioural data for AI supported early talent identification, the framework integrates principles of responsible AI, child centred data governance, explainable AI(XAI) to ensure that technological innovation remains aligned with developmental, educational, and societal priorities (Holmes et al., 2022; UNESCO, 2023). EMAI-T identifies emerging talent patterns by comparing multimodal profiles of the children with domain expert data. Therefore, ethical safeguards are embedded throughout the AI lifecycle to prevent harmful profiling, biased interpretation, and reductionist labelling of child abilities.

3.5.1. Child Consent and Assent

The proposed framework adopts a dual-layered consent process involving informed parental/institutional (art coaching centers) consent together with age- appropriate child assent. Parent and institutions are provided with transparent information regarding multimodal data collection, AI- supported analysis, data storage, and withdrawal rights. Child participants also receive developmentally appropriate explanations to support meaningful participation [79]

3.5.2. Privacy and Data Protection

Multimodal indicators include behavioural data and gestures of children, by considering the sensitivity, the EMAI-T framework adopts a privacy design architecture which is aligned with GDPR and trustworthy AI principles (European Parliament, 2016; Floridi & Cowls, 2022). Further multimodal children's data are protected through anonymization, encrypted storage, controlled access protocols and secured repositories.

To ensure that only information relevant to talent- related analysis is retained, data minimization principles are incorporated. Audit trails and secure deletion procedures strengthen accountability and long-term data governance. In addition, explainability mechanisms are incorporated to provide transparency regarding how behavioural indicators contribute to AI- generated interpretations.

3.5.3. Responsible AI and Explainable Talent Identification

Talent identification involves sensitive developmental interpretation, Explainable AI(XAI) mechanisms are incorporated to provide interpretable insights into the multimodal features influencing model outputs. Human educators, psychologists, and domain experts therefore remain central to interpretation and decision-making processes, ensuring that talent identification is treated as dynamic, context- sensitive, and developmentally evolving rather than fixed or deterministic. The proposed EMAI-T framework frames AI as an assistive educational tool rather than an autonomous decision-making system. Guided by the IEEE 7000 standards and UNESCO recommendations on trustworthy AI, the framework emphasizes transparency, explainability, accountability, and human oversight (IEEE, 2012; UNESCO, 2023).

The proposed framework establishes a proactive and theoretically grounded foundation for responsible multimodal talent identification in early childhood contexts. By embedding safeguards related to consent, privacy and explainability at the blueprint stage, EMAI-T seeks to support trustworthy and socially responsible AI deployment in early talent discovery.

3.6 Expected Contribution of the Dataset Blueprint

The proposed Multimodal Artistic- Creative Expression Dataset blueprint defines a theory- aligned observational foundation for the development of the EMAI-T framework. The framework integrates multidisciplinary theoretical grounding, structured multimodal representation design, construct valid annotation principles, and embedded ethical governance. The blueprint also advances a scalable infrastructure for responsible AI- assisted talent identification research.

The dataset design is not positioned merely as a data repository, but as construct sensitive modelling ecosystem. The Dataset preserves interpretability across representation, fusion, and evaluation stages. Through its emphasis on multimodal richness, cross-modal alignment, and explainable architecture readiness, the blueprint contributes methodological rigor to emerging efforts in strength-based AI systems for early childhood education.

The subsequent section defines the architectural and explainability principles through which these multimodal representations are computationally instantiated within the EMAI-T framework.

4. EMAI -T Framework Architecture and Explainability Mechanism

This section is consistent with Objective 3 and defines the architectural and explainability principles of the explainable Multimodal AI for talent identification (EMAI -T) framework. The proposed architecture operationalizes the theoretical constructs articulated in Section 2 and the dataset design blueprint described in Section 3 into a structured, interpretable multimodal AI system. The EMAI- T architecture is designed around transparency, construct alignment, and traceability from raw multimodal input to stakeholder-interpretable output.

4.1 Overview of the EMAI-T Design

The proposed EMAI-T framework adopts a layered architecture integrating multimodal representation learning with explainability mechanism. Multimodal learning has been shown to enhance representational richness by leveraging complementary sensory and behavioural stream. EMAI-T incorporates three conceptual layers which are built upon these paradigm

1) Multimodal Input and Encoding Layer

This layer processes synchronized textual, visual and behavioural signals.

2) Latent Representation and Fusion Layer

This layer integrates modality specific embeddings within a shared representational space.

3) Explainable Inference and output Layer

This layer generates interpretable predictions through embedded and post-hoc explainability techniques.

The layered structure establishes a traceable inference pathway aligned with principles of trustworthy AI [12] ensuring that model outputs remain accountable and pedagogically interpretable.

4.2 Multimodal Input and Encoding Layer

The multimodal input layer process synchronized multimodal signals derived from the dataset blueprint described in Section 3. Each modality is processed through domain-appropriate representation mechanisms.

Textual stream

Textual data such as self-reported responses to specifically designed questionnaires used to assess creative artistic behaviour in children. Domain experts evaluated these responses, and their assessments served as the reference labels. After basic preprocessing, the textual data will be encoded using a pretrained NLP based language model to generate semantic embeddings. These embedding will be combined with features from other modalities.

Visual Stream

Visual artifacts such as final drawings are encoded using convolutional neural networks (CNNs) or vision transformer architectures. These have demonstrated effectiveness in extracting hierarchical spatial representation. Feature representations may capture compositional complexity, stroke variability, colour distribution, and spatial organization attributes theoretically linked to visual-spatial intelligence.

Behavioural Stream

Drawing behaviour, Stroke dynamics, symmetry, colour variability gaze patterns, and gesture dynamics are encoded using affect recognition and pose estimation frameworks. These representations support modelling of engagement, motivation, and emotional resonance constructs aligned with intrapersonal and developmental catalysts (Gagné, 2015).

Temporal synchronization across modalities enables the formation of unified multimodal tensors, facilitating downstream fusion and cross-modal relational modelling.

4.3 Latent Representation and Fusion Strategy

Contemporary multimodal learning research identifies multiple fusion strategies, including early fusion, late fusion, and transformer-based cross-attention architectures (Baltrušaitis et al., 2019; Vaswani et al., 2017). Two complementary strategies are proposed within the conceptual design of EMAI-T

1) Early Fusion Pathway

In the early fusion pathway, modality-specific embeddings are integrated within a shared latent space using transformer-based encoders. This enables modelling of inter-modal dependencies (e.g., synchrony between gestural intensity and tonal variation).

2) Late Fusion Pathway

In the late fusion pathway modality-specific outputs are aggregated through weighted attention mechanisms, allowing adaptive emphasis on the most informative modalities.

4.4 Explainable Inference and Interpretability Layer

Explainability constitutes a foundational design principle of the EMAI-T architecture In, Interpretability is essential for accountability and stakeholder trust in high-stakes educational contexts (Holmes et al., 2022). The framework integrates intrinsic and post-hoc explainability techniques including Attention Visualization and Saliency Mapping and Model-Agnostic Explanations. Gradient-based techniques provide insights into feature importance within deep architectures. SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016) generate local and global feature attribution explanations. These mechanisms enable the translation of latent multimodal representations into human-readable insights. EMAI-T bridges computational reasoning with pedagogical interpretation, supporting a transparent-by-design modelling paradigm by embedding explainability at multiple levels

4.5 Ethical, Technical, and Practical Design Considerations

Within the architectural logic, ethical readiness is embedded. Responsible AI frameworks emphasize privacy preservation, secure data storage, anonymization and accountability mechanisms (Holmes et al., 2022). The EMAI-T design incorporates privacy-preserving extensions, such as federated learning or differential privacy, to mitigate sensitive data exposure and modular scalability, enabling extension to additional talent domains without restructuring core components.

The framework operationalizes ethical AI principles alongside computational robustness by integrating these mechanisms at the architectural level

5. Practical and Theoretical Implications

Developing AI systems to identify creative artistic behaviour in children requires careful consideration of ethical responsibility, pedagogical utility, and theoretical advancement. This section outlines the ethical foundations, practical implications for educational settings, and broader interdisciplinary contributions in consistent with the human-centred and strength-based orientation of the EMAI-T framework,

5.1 Ethical Foundations of the EMAI-T Framework

The integration of artificial intelligence into child-centred educational contexts incorporates ethical safeguards, particularly given the sensitivity of developmental and behavioural data (Holmes et al., 2022). The EMAI-T framework adopts an embedding transparency and privacy principles directly into its architectural logic.

Multimodal data collected from the children aged 8-12 including visual artifacts, Textual pattern, behaviour signals constitute sensitive developmental information that demands robust governance (UNICEF, 2021). The framework aligns with international ethical standards, including the UN Convention on the Rights of the Child (United Nations, 1989), UNICEF's Policy Guidance on AI for Children (2021), the General Data Protection Regulation (European Parliament, 2016), and IEEE's Ethically Aligned Design framework (IEEE, 2019).

The interrelated ethical pillars underpin the framework are

(1) Transparency and Explainability.

Explainability supports accountability and enables stakeholders to interrogate algorithmic reasoning in educational settings, rather than accept opaque outputs (Holmes et al., 2022). Interpretability is embedded through attention visualization and model-agnostic explanation techniques such as SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro et al., 2016).

(2) Privacy and Informed Consent.

The standard data governance procedures are structured around informed parental consent, child assent, anonymization protocols, and secure data handling in compliance with GDPR standards (European Parliament, 2016). These mechanisms recognize children's rights to data protection and participatory dignity (UNICEF, 2021). By embedding ethical governance at procedural, algorithmic, and institutional levels, EMAI-T aligns with emerging frameworks for trustworthy AI in education (Holmes et al., 2022).

5.2 Practical Implications for Educational Practice

EMAI-T has implications for pedagogical practice and institutional readiness within early childhood education beyond the architectural contributions

(1) Supporting Educator Insight and Differentiated Instruction

Educators often encounter challenges in identifying underlying talents that are not reflected in standardized academic measures (Subotnik et al., 2011). The framework offers decision-support tools that may inform differentiated instruction and personalized learning strategies by providing interpretable visualizations of creative indicators such as expressive diversity, compositional complexity, and affective engagement

Moreover, the system is positioned as an augmentation tool rather than a replacement for professional judgment, consistent with human-centred AI principles (Shneiderman, 2020).

(2) Advancing Strength-Based Assessment Paradigms

Rather than recognition of potential strengths, educational assessment has historically emphasized remediation of deficits (Sternberg, 2005). EMAI-T contributes to a conceptual shift toward strength-based profiling by foregrounding creative capacities and expressive individuality in child.

Including Multiple Intelligences-informed instruction (Gardner, 1999) and child-centred learning approaches, this orientation aligns with constructivist and holistic pedagogical philosophies. The system supports early identification and nurturing of artistic potential rather than reactive correction of academic gaps by operationalizing creative indicators within an interpretable AI framework.

5.3 Theoretical and Interdisciplinary Contributions

The EMAI-T framework advances at the intersection of educational psychology and artificial intelligence through three primary contributions.

(1) interdisciplinary operationalization of psychological theory

The framework translates constructs from Gardner's Multiple Intelligences theory (Gardner, 1983, 1999) and Gagné's DMGT (Gagné, 2015) into computationally interpretable representations. This integration moves beyond superficial citation of theory by embedding construct validity within model design.

(2) conceptual reframing of AI-driven talent identification.

EMAI-T contributes to reimagine the role of AI in education as a facilitator of potential recognition rather than deficit detection by shifting emphasis from diagnostic classification to developmental discovery,

(3) interpretability as an epistemological principle.

Explainability is positioned as a foundational knowledge-based requirement for educational AI. By incorporating interpretability, transparency becomes integral to know how knowledge development is constructed, interpreted, and acted upon.

These contributions extend the discourse on explainable AI in education from performance optimization toward psychologically grounded and ethically responsible design.

6. Limitations

The present framework does not include empirical validation and performance evaluation as it is conceptual and methodological blueprint. Subsequent studies will implement the dataset, construct expert-reference embeddings and empirically assess similarity-based modelling outcomes.

7. Conclusion and Future Directions

The proposed study has presented the theoretical and methodological foundation of the Explainable Multimodal AI for Talent Identification (EMAI-T) framework. EMAI -T is a multidisciplinary design that integrates developmental psychology, giftedness theory, and explainable artificial intelligence to support identification of artistic-creative potential in children. EMAI-T framework is positioned against performance-oriented AI systems in education. EMAI-T advances a strength-based and transparency-centered paradigm that emphasizes psychological validity, interpretability

The paper addressed three primary objectives.

Objective 1 defines a structured design blueprint for a multimodal reference dataset. Reference dataset integrates Textual visual, and behavioural indicators of creativity across various age groups. The dataset design establishes a theoretically aligned empirical foundation for future model training and benchmarking.

Objective 2 develops an interdisciplinary conceptual framework by operationalizing constructs from Gardner's Multiple Intelligences theory and Gagné's Differentiated Model of Giftedness and Talent (DMGT) into computationally interpretable representations. This translation from theory to model architecture strengthens construct validity and ensures alignment between psychological meaning and algorithmic inference.

Objective 3 proposes the multimodal, explainable AI EMAI-T architectural framework. The EMAI-T embeds interpretability mechanisms such as attention-based visualization and model-agnostic explanation techniques such as SHAP and LIME. The framework also establishes traceability between raw multimodal inputs and stakeholder-interpretable outputs. This design advances AI in education beyond predictive accuracy toward accountable and human-centred modelling.

These contributions position EMAI-T as a conceptual and methodological foundation for responsible AI-driven talent identification in children. The framework is designed to augment educator insight through transparent, theory-informed computational analysis rather than replacing professional judgment. In doing so, it contributes to emerging discourse on trustworthy AI in education and interdisciplinary integration between psychological theory and machine learning design.

Future research will focus on empirical instantiation and validation of the framework (Objective 4). Planned pilot studies involving children aged 8–12 years, educators, and caregivers will examine interpretability, usability, fairness, and predictive performance in applied educational settings.

The EMAI-T framework provides a structured pathway for aligning multimodal AI modelling with developmental theory, and educational interpretability. This work establishes a foundation for future empirical research and responsible deployment of AI systems aimed at recognizing and nurturing early human potential by integrating construct validity, multimodal representation, and embedded explainability

References

1. Gardner, H. (2011). Frames of mind: The theory of multiple intelligences. Basic books.
2. Kaufman, J. C., & Sternberg, R. J. (2023). The Cambridge handbook of creativity (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781009119464
3. Sternberg, R. J. (2005). The WICS Model of Giftedness, u: Sternberg, Robert J., Janet E. Davidson (ed.), Conceptions of giftedness.
4. Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542–570. https://doi.org/10.1111/ejed.12533
5. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(39), 1–27. https://doi.org/10.1186/s41239-019-0171-0
6. Jose, B., Cherian, J., Verghis, A. M., & Varghise, S. M. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. Frontiers in Psychology, 16, 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

7. Li, X., & Zhao, Y. (2022). Bridging AI and developmental psychology: Conceptual frameworks for intelligent learning environments. Computers & Education, 184, 104517. https://doi.org/10.1016/j.compedu.2022.104517
8. Guerrero-Sosa, J. D. T., Romero, F. P., et al. (2025). A comprehensive review of multimodal analysis in education. Applied Sciences, 15(11), 5896. https://doi.org/10.3390/app15115896
9. Stanja, J., Gritz, W., Krugel, J., Hoppe, A., & Dannemann, S. (2023). Formative assessment strategies for students' conceptions—The potential of learning analytics. British Journal of Educational Technology, 54(2), 355–370. https://doi.org/10.1111/bjet.13288
10. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
11. Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of explainable AI techniques in healthcare. Sensors, 23(2), 634. https://doi.org/10.3390/s23020634
12. Whalen, J., & Mouza, C. (2023). ChatGPT: Challenges, opportunities, and implications for teacher education. Contemporary Issues in Technology and Teacher Education, 23(1). https://www.learntechlib.org/p/222408/
13. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32, 504–526. https://doi.org/10.1007/s40593-021-00239-1
14. Guerrero-Sosa, J. D. T., Romero, F. P., Menéndez-Domínguez, V. H., Serrano-Guerrero, J., Montoro-Montarroso, A., & Olivas, J. A. (2025). A comprehensive review of multimodal analysis in education. Applied Sciences, 15(11), 5896. https://doi.org/10.3390/app15115896
15. Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. Creativity Research Journal, 24(1), 92–96. https://doi.org/10.1080/10400419.2012.650092
16. K. Ahmad et al., "Data-Driven Artificial Intelligence in Education: A Comprehensive Review," in IEEE Transactions on Learning Technologies, vol. 17, pp. 12-31, 2024, https://doi.org/10.1109/TLT.2023.3314610
17. Liu, R., & Glave, C. Z. (2026). Justice, Equity, Diversity, and Inclusion in Artificial Intelligence and Childhood Education: A Critical Systematic Review. Journal of Research in Childhood Education, 40(1), 5–23. https://doi.org/10.1080/02568543.2025.2578526
18. Paolucci, C., Vancini, S., Bex II, R. T., Cavanaugh, C., Salama, C., & de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for K-12 education. Heliyon, 10(4). DOI: 10.1016/j.heliyon.2024.e25767
19. 18] T. Baltrušaitis, C. Ahuja and L. -P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423-443, 1 Feb. 2019. DOI: 10.1109/TPAMI.2018.2798607
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. DOI link: https://doi.org/10.48550/arXiv.1706.03762
21. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30. https://doi.org/10.48550/arXiv.1705.07874
22. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778
23. Sternberg, R. J. (2005). The WICS Model of Giftedness, u: Sternberg, Robert J., Janet E. Davidson (ed.), Conceptions of giftedness.
24. Winner, E. (1997). Gifted Children: Myths and Realities.
25. Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: A proposed direction forward based on psychological science. Psychological science in the public interest, 12(1), 3-54. https://doi.org/10.1177/1529100611418056
26. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education promises and implications for teaching and learning. Center for Curriculum Redesign. https://curriculumredesign.org/our-work/artificial-intelligence-in-education/
27. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144) https://doi.org/10.1145/2939672.2939778
28. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30. Google Scholar□
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. GoogleScholar□
30. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504-526. https://doi.org/10.1007/s40593-021-00239-1
31. Benvenuti, M., Cangelosi, A., & Weinberger, A. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. Computers in Human Behavior, 147, 107859. https://doi.org/10.1016/j.chb.2023.107859
32. Borenstein, J., Herkert, J. R., & Miller, K. W. (2020). The ethics of artificial intelligence: Ethical theory and AI in the age of autonomy. Science and Engineering Ethics, 26(5), 1–23. https://doi.org/10.1007/s11948-019-00146-3

33. Cohn, C., Snyder, C., Fonteles, J. H., & TS, A. (2025). A multimodal approach to support teacher, researcher, and AI collaboration in STEM+C learning environments. British Journal of Educational Technology, 56(1), 55–72. https://doi.org/10.1111/bjet.13455
34. Crescenzi-Lanna, L. (2023). The reciprocal value of artificial and human intelligence in early childhood education: A systematic review. Journal of Research on Technology in Education, 55(4), 567–589. https://doi.org/10.1080/15391523.2023.2231124
35. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1
36. Gagné, F. (2004). Transforming gifts into talents: The DMGT as a developmental theory. High Ability Studies, 15(2), 119–147. https://doi.org/10.1080/1359813042000314682
37. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. IEEE Transactions on Neural Networks and Learning Systems, 31(10), 3459–3473. https://doi.org/10.1109/TNNLS.2018.2892568
38. Giovanola, B., & Granata, P. (2024). Ethics for human-centered education in the age of AI. Edward Elgar Publishing. https://doi.org/10.4337/9781035304018
39. Heeg, D. M., & Avraamidou, L. (2025). Young children's understanding of artificial intelligence. Education and Information Technologies, 30(2), 247–264. https://doi.org/10.1007/s10639-024-12167-2
40. Holmes, W., Bialik, M., & Fadel, C. (2022). Artificial intelligence in education: Promises and implications for teaching and learning. Computers & Education, 190, 104598. https://doi.org/10.1016/j.compedu.2022.104598
41. Kaufman, J. C., & Sternberg, R. J. (2023). The Cambridge handbook of creativity (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781009119464
42. Lee, G., Shi, L., Latif, E., & Gao, Y. (2025). Multimodality of AI for education: Towards artificial general intelligence. IEEE Transactions on Learning Technologies, 18(1), 121–135. https://doi.org/10.1109/TLT.2025.3356724
43. 42] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774. https://doi.org/10.48550/arXiv.1705.07874
44. Maity, S., & Deroy, A. (2024). Human-centric explainable AI in education. arXiv preprint. https://doi.org/10.48550/arXiv.2410.19822
45. Niu, T., Liu, T., Luo, Y. T., Pang, P. C. I., Huang, S., & Xiang, A. (2025). Decoding student cognitive abilities: A comparative study of explainable AI algorithms in educational data mining. Scientific Reports, 15(1), 4512. https://doi.org/10.1038/s41598-025-31412-4
46. Ozmen Garibay, O., Winslow, B., & Andolina, S. (2023). Six human-centered artificial intelligence grand challenges. International Journal of Human–Computer Interaction, 39(6), 1075–1093. https://doi.org/10.1080/10447318.2023.2176501
47. Pisoni, G., Díaz-Rodríguez, N., Gijlers, H., & Tonolli, L. (2021). Human-centered artificial intelligence for designing accessible cultural heritage. Applied Sciences, 11(2), 870. https://doi.org/10.3390/app11020870
48. Porayska-Pomsta, K., & Rajendran, G. (2019). Accountability in human and artificial intelligence decision-making as the basis for diversity and educational inclusion. In Artificial Intelligence and Inclusive Education (pp. 45–62). Springer. https://doi.org/10.1007/978-3-030-23204-7_3
49. Renz, A., & Vladova, G. (2021). Reinvigorating the discourse on human-centered artificial intelligence in educational technologies. Technology Innovation Management Review, 11(5), 38–[48] https://doi.org/10.22215/timreview/1443
50. Su, J., & Zhong, Y. (2022). Artificial Intelligence (AI) in early childhood education: Curriculum design and future directions. Computers & Education: Artificial Intelligence, 3, 100056. https://doi.org/10.1016/j.caeai.2022.100056
51. Vincent-Lancrin, S. (2020). Trustworthy artificial intelligence (AI) in education: Promises and challenges. OECD Education Working Papers. https://doi.org/10.1787/a6c90fa8-en
52. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence in education. Computers & Education, 146, 103751. https://doi.org/10.1016/j.compedu.2019.103751
53. Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. Humanities and Social Sciences Communications, 12(1), 201. https://doi.org/10.1057/s41599-025-01854-x
54. Mohammadi, Mehrnoush, et al. “Artificial Intelligence in Multimodal Learning Analytics: A Systematic Literature Review.” Computers and Education: Artificial Intelligence, vol. 8, Jun. 2025, p. 100426. DOI.org (Crossref), https://doi.org/10.1016/j.caeai.2025.100426
55. N. Rodis, C. Sardianos, P. Radoglou-Grammatikis, P. Sarigiannidis, I. Varlamis and G. T. Papadopoulos, "Multimodal Explainable Artificial Intelligence: A Comprehensive Review of Methodological Advances and Future Research Directions," in IEEE Access, vol. 12, pp. 159794-159820, 2024, https://doi.org/10.1109/ACCESS.2024.3467062
56. . Aravind, Angalakuduru, et al. “Adaptive AI-Based Personalized Learning for Accelerated Vocabulary and Syntax Mastery in Young English Learners.” International Journal of Advanced Computer Science and Applications (Ijacsa), vol. 16, no. 4, Mar. 2025. thesai.org, https://doi.org/10.14569/IJACSA.2025.0160467
57. Mawalim, Candy Olivia, et al. “Beyond Accuracy: Multimodal Modeling of Structured Speaking Skill Indices in Young Adolescents.” Computers and Education: Artificial Intelligence, vol. 8, Jun. 2025, p. 100386. DOI.org (Crossref), https://doi.org/10.1016/j.caeai.2025.100386

58. Prova, Nuzhat, Explainable AI-Powered Multimodal Fusion Framework for EEG-Based Autism Spectrum Disorder Classification (January 28, 2025). Available at SSRN: https://ssrn.com/abstract=5114986 or http://dx.doi.org/10.2139/ssrn.5114986
59. Shah, Uzair, et al. "ArtInsight: A Multimodal AI Framework for Interpreting Children's Drawings and Enhancing Emotional Understanding." Studies in Health Technology and Informatics, edited by Elisavet Andrikopoulou et al., IOS Press, 2025. DOI.org (Crossref), https://doi.org/10.3233/SHTI250471
60. Guerrero-Sosa, Jared D. T., et al. "A Multimodal Framework for Explainable Evaluation of Soft Skills in Educational Environments." arXiv:2505.01794, arXiv, 3 May 2025. arXiv.org, https://doi.org/10.48550/arXiv.2505.01794
61. Zhu, F. L., Wang, S. H., Liu, W. B., Zhu, H. L., Li, M., & Zou, X. B. (2023). A multimodal machine learning system in early screening for toddlers with autism spectrum disorders based on the response to name. Frontiers in psychiatry, 14, 1039293. https://doi.org/10.3389/fpsyt.2023.1039293
62. Khosravi, Hassan, et al. "Explainable Artificial Intelligence in Education." Computers and Education: Artificial Intelligence, vol. 3, 2022, p. 100074. DOI.org (Crossref), https://doi.org/10.1016/j.caeai.2022.100074
63. Aylward, B. S., Abbas, H., Taraman, S., Salomon, C., Gal-Szabo, D., Kraft, C., Ehwerhemuepha, L., Chang, A., & Wall, D. P. (2023). An Introduction to Artificial Intelligence in Developmental and Behavioral Pediatrics. Journal of developmental and behavioral pediatrics : JDBP, 44(2), e126–e134. https://doi.org/10.1097/DBP.0000000000001149
64. Agrawal, R., & Agrawal, R. (2025). Explainable AI in early autism detection: a literature review of interpretable machine learning approaches. Discover mental health, 5(1), 98. https://doi.org/10.1007/s44192-025-00232-3
65. Khosravi, Hassan, et al. "Explainable Artificial Intelligence in Education." Computers and Education: Artificial Intelligence, vol. 3, 2022, p. 100074. DOI.org (Crossref), https://doi.org/10.1016/j.caeai.2022.100074
66. Sun, Shilin, et al. "A Review of Multimodal Explainable Artificial Intelligence: Past, Present and Future." arXiv, 2024. DOI.org (Datacite), https://doi.org/10.48550/ARXIV.2412.14056
67. Shaheen, Nour, et al. "Appraising Systematic Reviews: A Comprehensive Guide to Ensuring Validity and Reliability." Frontiers in Research Metrics and Analytics, vol. 8, Dec. 2023, p. 1268045. DOI.org (Crossref), https://doi.org/10.3389/frma.2023.1268045
68. Mancin, S., Sguanci, M., Andreoli, D., Soekeland, F., Anastasi, G., Piredda, M., & De Marinis, M. G. (2023). Systematic review of clinical practice guidelines and systematic reviews: A method for conducting comprehensive analysis. MethodsX, 12, 102532. https://doi.org/10.1016/j.mex.2023.10253
69. . Benson, F. N., Chelangat, D., Brink, W., Mwangala, P. N., Waljee, A. K., Moyer, C. A., & Abubakar, A. (2025). Application of machine learning in early childhood development research: a scoping review. BMJ open, 15(8), e100358. https://doi.org/10.1136/bmjopen-2025-100358
70. Reinhart, Lisa, et al. "Artificial Intelligence in Child Development Monitoring: A Systematic Review on Usage, Outcomes and Acceptance." Intelligence-Based Medicine, vol. 9, 2024, p. 100134. DOI.org (Crossref), https://doi.org/10.1016/j.ibmed.2024.100134.
71. Cui, Kaijie, et al. "Development of an Artificial Intelligence-Based Multimodal Model for Assisting in the Diagnosis of Necrotizing Enterocolitis in Newborns: A Retrospective Study." Frontiers in Pediatrics, vol. 12, May 2024, p. 1388320. DOI.org (Crossref), https://doi.org/10.3389/fped.2024.1388320.
72. Sadeghi, Zahra, et al. "A Review of Explainable Artificial Intelligence in Healthcare." Computers and Electrical Engineering, vol. 118, Aug. 2024, p. 109370. DOI.org (Crossref), https://doi.org/10.1016/j.compeleceng.2024.109370
73. . Muhammad, Dost, and Malika Bendechache. "Unveiling the Black Box: A Systematic Review of Explainable Artificial Intelligence in Medical Image Analysis." Computational and Structural Biotechnology Journal, vol. 24, Dec. 2024, pp. 542–60. DOI.org (Crossref), https://doi.org/10.1016/j.csbj.2024.08.005
74. . Bae, S., Hong, J., Ha, S. et al. Multimodal AI for risk stratification in autism spectrum disorder: integrating voice and screening tools. npj Digit. Med. 8, 538 (2025). https://doi.org/10.1038/s41746-025-01914-6
75. Cohn, Clayton, et al. "A Multimodal Approach to Support Teacher, Researcher and AI Collaboration in STEM +C Learning Environments." British Journal of Educational Technology, vol. 56, no. 2, Mar. 2025, pp. 595620. DOI.org (Crossref), https://doi.org/10.1111/bjet.13518.
76. Brankovic, A., Cook, D., Rahman, J. et al. Clinician-informed XAI evaluation checklist with metrics (CLIX-M) for AI-powered clinical decision support systems. npj Digit. Med. 8, 364 (2025). https://doi.org/10.1038/s41746-02501764-2