

Modeling an AI-Powered Vietnamese Folk Paintings as an Inclusive Educational Platform for Children with Disabilities

A Human-Centred Multimodal Architecture

CAO Phương Thao PHAM Phương Linh NGUYEN Le Quoc Bao

Data scientist, France

Intern in AI

Inter in AI

Abstract. Most learning materials assume fluent speech, typical vision and hearing, fine motor control and a single way of responding. Many children with disabilities do not meet one or more of these assumptions and are shut out as a result. Vietnamese folk paintings from the Đông Hồ, Hàng Trống, Kim Hoàng and Sinh traditions show recognizable scenes of animals, family life, work and festivals, and with the right support they suit vocabulary, sequencing, storytelling and cultural discussion. We model an AI-powered platform that connects these artworks to accessible learning activities. Following a design science approach, we analyze stakeholder needs, derive design principles and requirements, and specify a conceptual model of activity selection together with a layered architecture. Five functional layers (user interaction, AI services, knowledge, data management and cloud-edge deployment) sit beside a governance column that spans all of them. The AI services cover visual analysis of artworks, optional speech input and output, language generation grounded in reviewed cultural records, transparent activity recommendation, and an adaptive policy that is disabled by default. A cultural knowledge graph keeps what is visible in an image apart from what commentators say it means. Learner profiles are minimal and tied to specific tasks, and educators can inspect, override or reset every recommendation. We formalize the selection and adaptation rules, describe the artwork record schema, set out a privacy and security architecture and discuss the trade-offs involved. The platform is not a medical device. No dataset was collected and no experiment is reported; the paper provides an architecture that can be built and then evaluated with educators, families and learners.

Keywords: inclusive education; Vietnamese folk paintings; cultural heritage; multimodal AI; system architecture; assistive technology; knowledge graph; human-centred AI.

1. Introduction

Inclusive education means teaching all children together and adjusting materials, interaction and assessment so that differences in communication, sensory, cognitive and motor abilities do not exclude anyone [1, 2]. Digital tools can help. An application can repeat an instruction as often as needed, read text aloud, accept a tap instead of a spoken answer and let the learner set the pace. The same application becomes another barrier as soon as it demands clear speech, precise pointing, sustained attention or one fixed route through a lesson. In low- and middle-income countries the gap in access to assistive products is especially wide [3].

Take a Đông Hồ woodblock print of a fat pig, printed in bold outlines on paper coated with scallop shell. Most Vietnamese children would recognize the image at once, even though the craft that produces it is close to disappearing: UNESCO placed Đông Hồ printmaking on its List of Intangible Cultural Heritage in Need of Urgent Safeguarding in 2025 because only a few families still practise it [4]. Familiar material is a sensible starting point for language and concept work. Folk paintings from the Đông Hồ, Hàng Trống, Kim Hoàng and Sinh

traditions, with their scenes of animals, family life, harvests and festivals, lend themselves to naming objects, building vocabulary, ordering events and telling stories, often with little adaptation.

Their value for a particular child still depends on the child and the picture. Pictures differ in visual complexity and in how much background knowledge they require, symbols are read differently across regions, and a devotional Sinh print carries connotations that a cheerful Đông Hồ animal scene does not. A teacher still has to choose and adapt.

AI could take over part of that work by finding objects in an image, reading descriptions aloud, accepting spoken answers, drafting questions and suggesting what to try next. The risks are serious. Wrong object labels mislead, invented cultural facts distort a tradition that is already fragile, speech recognition breaks down on atypical speech, emotion inference from faces is unreliable, and children's data can accumulate well beyond what any task needs [5–7].

We model a platform that uses AI to make folk paintings accessible as learning material while teachers, therapists and families stay in charge. The paintings are starting points for learning activities and are not presented as therapy. Every AI output is a suggestion that a responsible adult can accept, edit or reject; when the users are children with disabilities, the line between an assistive suggestion and an automated decision matters a great deal. No dataset was collected and no experiment is reported.

Four research questions guide the work. RQ1: how can Vietnamese folk paintings be represented as rights-aware, accessible learning resources? RQ2: how can multimodal AI support individualized activities without replacing professional judgement? RQ3: what architecture and data model can deliver these functions while limiting the collection and exposure of children's data? RQ4: what governance mechanisms does such a platform need at each layer?

The paper analyzes stakeholders and derives requirements for culturally grounded inclusive learning, proposes a formal model of how activities are filtered, ranked and adapted (Section 4.2), and specifies a layered architecture with a governance column that runs across all five layers. A provenance-aware artwork schema and a cultural knowledge graph keep observable content apart from sourced interpretation, and the privacy design limits what the platform stores about children.

2. Background

2.1 Vietnamese folk painting traditions

Vietnamese woodblock printing developed in several village and urban centres, each with its own technique and purpose. Maurice Durand's study of Vietnamese popular imagery documented two of them in detail: Đông Hồ, with its flat printed colour, and Hàng Trống, where colour is applied with a brush for finer gradation [8].

Đồng Hồ prints come from a village in Bắc Ninh province. Artisans carve each design into woodblocks and print the colours one after another by hand on paper coated with powdered scallop shell, with the black outline printed last; the pigments come from natural materials. Subjects include worship, history, daily life and landscapes, and the prints are closely tied to the Lunar New Year, the Mid-Autumn Festival and ancestral rites. Among the best-known subjects are pigs, roosters, buffaloes, children at play and *Đám cưới chuột* (The Rats' Wedding). Only a few families still practise the craft, which is why UNESCO inscribed it on the Urgent Safeguarding List in 2025 [4].

Hàng Trống pictures come from the old commercial quarter of Hà Nội. They combine a printed outline with colour applied by brush, which allows finer gradation, and they include both decorative and devotional subjects [8].

Kim Hoàng, a village on the outskirts of Hà Nội, uses another process. Outlines are printed on red or yellow paper and the details are painted by hand, and roosters, pigs and domestic scenes expressing wishes for prosperity dominate. A flood in the 1940s washed away the village's woodblocks, and for decades the genre was no longer produced. A revival began in 2015, when a collector set out to recover surviving woodblocks and worked with craftsmen and local authorities to restart production [9].

Sinh pictures are made in a village near Huế for worship. They are used in ceremonies, Tết offerings and rituals, and are traditionally burned afterwards. Artisans print them with woodblocks on paper coated with ground seashell and colour them by hand [10]. Their spiritual function is something any educational use has to respect.

These differences matter for a learning platform. Devotional and ritual images carry meanings that may make them unsuitable for casual classroom use, so not every image belongs in the collection. Some traditions survive mainly in museum collections and private holdings, which means reproduction rights have to be checked image by image, and sources disagree about the social commentary in particular scenes. Presenting one reading as settled fact would mislead.

2.2 Inclusive education and accessibility

Universal Design for Learning (UDL) asks that a lesson work for a wide range of learners from the start instead of being retrofitted later, by offering multiple means of engagement, representation and action [11]. Version 3.0 (2024) also asks designers to address barriers rooted in bias and exclusion. WCAG 2.2 turns some of these ideas into testable criteria for digital interfaces, among them text alternatives, keyboard operability, sufficient contrast, adjustable timing and adequate target sizes [12].

Many children with complex communication needs use augmentative and alternative communication (AAC), which includes symbol boards, speech-generating devices, signs and gestures. Light and McNaughton define communicative competence for AAC users as the ability to express needs, share information and take part in social interaction through whatever

modes work for the person [13]. A system that scores only spoken answers will misjudge many of these children.

Rights frameworks point the same way. The Convention on the Rights of Persons with Disabilities [2] and the Convention on the Rights of the Child [14] both call for inclusive participation, and in Viet Nam the Law on Persons with Disabilities and the Law on Children set out related obligations [15, 16].

2.3 AI for learners with disabilities

Speech recognition, intelligent tutoring, emotion recognition, social robots and learning analytics have all been tried in special education, and reviews report small samples, short studies and limited attention to ethics [17, 18]. Large language models add another tool. They can produce explanations, questions and stories at different reading levels within seconds, and they can also produce fluent text that is factually wrong, which makes them both useful and risky in education [6, 19]. Ethical frameworks for AI in education stress fairness, privacy, autonomy and pedagogical purpose [7], and UNICEF's guidance adds requirements on development, inclusion and explanation that apply to younger users [20].

Automatic emotion recognition needs particular caution. A large review found that facial movements do not reliably reveal emotional state across people, situations and cultures [5]. For children with disabilities, whose expressions, gaze and movements may differ from those in training populations, the risk of misreading is higher still.

2.4 AI for cultural heritage

Machine learning has made progress in classifying artworks, attributing styles, detecting objects and enriching archive metadata [21, 22], mostly with convolutional networks [23], Vision Transformers [24] and transformer-based detectors [25]. Stylized folk prints look very different from the photographs on which most of these models are pre-trained, so their accuracy on such prints cannot be assumed. Heritage institutions also need to model provenance, events and relationships, and ontologies such as the CIDOC Conceptual Reference Model [26] and knowledge graphs [27] offer ways to represent entities, relations and their sources.

2.5 Research gap

Work on accessible learning technology rarely deals with cultural grounding, and heritage digitization rarely considers learners with disabilities. Few published designs connect a curated folk-art collection to accessible activities while keeping human oversight and children's privacy central, and we are not aware of any that combines accessibility, cultural heritage and AI governance in one architecture; confirming this would require a systematic search, which is outside the scope of the paper. Our aim is narrower than closing that gap: to show how the three concerns can be joined in one platform that teachers control and that collects as little data as possible.

3. Design approach and requirements

3.1 Design science approach

Design science research produces and evaluates artefacts that address a defined problem [28]. Peffers et al. describe its stages as problem identification, objectives, design, demonstration, evaluation and communication [29]. This paper covers the first three; demonstration and evaluation will follow once an implementation exists.

3.2 Stakeholders

Six groups have a stake in the platform (Table 1). Their needs are drawn from the literature reviewed above and from the roles the platform has to support. Preparation time is a practical constraint for teachers in special education, and several design decisions below aim to keep it low.

Table 1. Stakeholders and their needs.

Stakeholder	Main needs	Design implication
Child learners with diverse needs	Several ways to perceive and respond; control over pace; no penalty for stopping	Multimodal interface, AAC support, pause and skip at any time
Teachers	Activities matched to objectives; little setup time; clear reasons for suggestions	Dashboard with editable recommendations and explanations
Therapists and specialists	Alignment with individual plans; no conflict with clinical work	Activities framed as practice chosen by professionals, not therapy
Parents and caregivers	Simple home activities; understandable summaries	Caregiver app with limited, authorized information
Cultural institutions and artisans	Correct attribution, respected rights and meanings	Provenance and rights fields; cultural review of interpretations
Platform operators	Maintainable, secure, affordable system	Modular services, cloud-edge split, versioned models

3.3 Design principles

Seven principles guide the design. Some are standard good practice, and others respond to risks specific to children with disabilities and to cultural content.

Accessibility first. Where practical, every activity offers more than one way to receive information and more than one way to respond, following UDL and WCAG [11, 12].

Cultural grounding. Descriptions and stories are generated only from reviewed metadata and sourced context, and interpretive claims carry their source, since presenting one scholar's reading as uncontested fact would mislead.

Proportional data use. The platform collects only what a declared purpose requires and deletes it on schedule.

Uncertainty awareness. When a model is unsure, it asks for clarification or offers choices instead of marking an answer wrong. For a child who may already doubt their abilities, being told a correct answer is wrong does more harm than having an incorrect answer accepted.

Human authority. Teachers, therapists and families keep control over goals, activities and any adaptation [30, 31]. The rest of the design depends on this principle. A system that overrode a teacher's judgement about a child with a disability would be making automated decisions in a domain where automation has no legitimate authority.

Adapting tasks, not labelling children. Adaptation changes what the activity demands (difficulty, format, length) and never assigns a fixed ability label. A diagnosis may inform a starting point but never determines content by itself, because two children with the same diagnosis often have very different strengths and needs in a learning task.

Auditability. Every recommendation, the evidence behind it and the adult's decision about it are recorded so that they can be reviewed later.

3.4 Scope boundaries

The platform is an educational tool and not a medical device. Blurring this boundary would expose it to regulatory obligations it cannot meet and could mislead families about what the technology is for. It does not diagnose disabilities or emotional states and does not prescribe therapy. Where we mention therapeutic use, we mean activities that a therapist chooses for practice within an individual plan. Any claim of clinical or educational benefit would require its own study under ethical approval.

3.5 Requirements

Table 2 lists the functional (FR) and non-functional (NFR) requirements and the parts of the architecture that address them.

Table 2. Requirements and their architectural responses.

ID	Requirement	Addressed in
FR1	Present each approved artwork with text, audio description and adjustable visual settings	Interaction layer (Sec. 6.1)
FR2	Accept responses by touch, keyboard, switch, symbols, drawing, speech or adult-supported input	Interaction layer (Sec. 6.1); speech service (Sec. 7.2)
FR3	Suggest activities that match an educator's objective and the learner's access settings	Recommendation service (Sec. 7.4); conceptual model (Sec. 4)
FR4	Explain every suggestion and let educators accept, edit or reject it	Dashboard (Sec. 6.2)
FR5	Generate questions, simplified descriptions and story prompts only from reviewed sources	Language service (Sec. 7.3); knowledge layer (Sec. 8)
FR6	Record provenance, rights and review status for each artwork and interpretation	Knowledge layer (Sec. 8.1, 8.2)
FR7	Produce draft progress summaries for review by a responsible professional	Dashboard (Sec. 6.2)
NFR1	Meet WCAG 2.2 level AA for all interfaces	Interaction layer
NFR2	Keep working, in a reduced form, when AI services or the network fail	Graceful degradation (Sec. 7.7); edge processing (Sec. 9.4)
NFR3	Minimize, pseudonymize and time-limit personal data	Data layer (Sec. 9); governance (Sec. 10)
NFR4	Restrict access to children's records by role	Security architecture (Sec. 10.3)
NFR5	Version models, prompts and content so that changes can be traced and reversed	Model repository (Sec. 9.3)
NFR6	Avoid inference of emotions or sensitive traits by default	AI services (Sec. 7.6)

4. Conceptual model

4.1 Inputs, process and outputs

Figure 1 traces how an activity is selected and presented. Five inputs feed the process: an approved artwork with verified metadata, a learning objective set by an educator, the learner's access and communication settings, recent task evidence that an adult has validated, and constraints such as session length or content to avoid. The platform filters its activity library, ranks the eligible activities and shows the top suggestions to the educator, who decides. The chosen activity is then presented accessibly, and the learner responds in any supported mode. A minimal event record is stored and, once validated, updates a task-specific profile that

informs the next cycle. An adult can pause, change the format or switch adaptation off at any point.

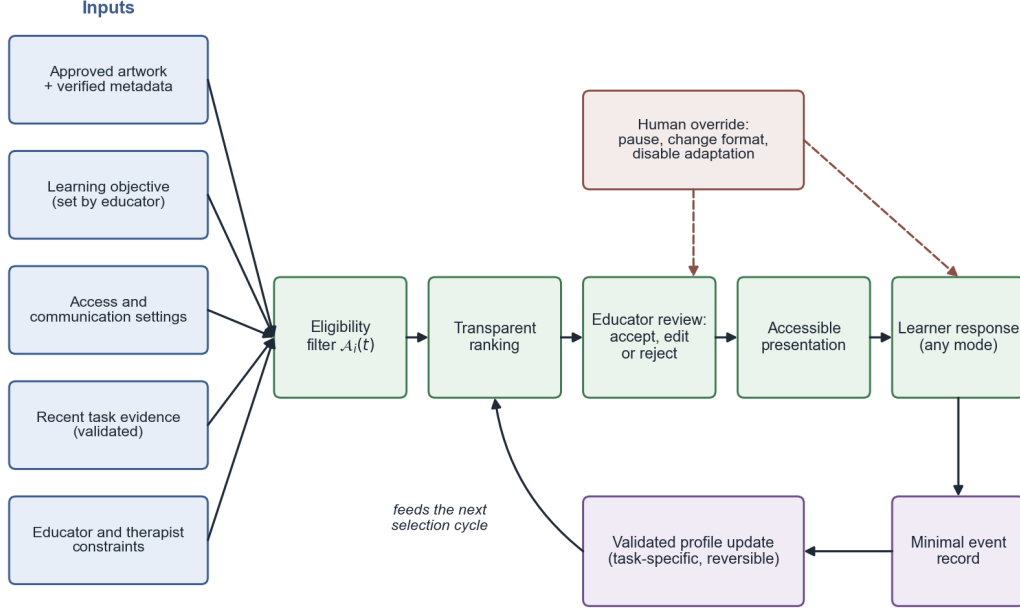


Figure 1. Conceptual model of activity selection. Inputs on the left are filtered and ranked; an educator reviews the suggestion before it reaches the learner, and validated evidence feeds the next cycle.

4.2 Formal model of activity selection

Let $\mathcal{A}$ be the library of approved activity templates, each instantiated for a specific artwork; templates and content still awaiting review are not part of $\mathcal{A}$. For learner i at time t , let $G_i(t)$ be the set of current objectives, M_i the set of supported input and output modes, L_i the maximum session length and X_i the content the educator has excluded. An activity a is eligible if its required modes are supported, it serves a current objective, it fits the time available and it has not been excluded:

$$\mathcal{A}_i(t) = \{a \in \mathcal{A} : \text{modes}(a) \subseteq M_i, \text{goal}(a) \in G_i(t), \text{dur}(a) \leq L_i, a \notin X_i\} \quad (1)$$

Eligible activities are ranked by a transparent weighted score:

$$S(a) = w_1 \text{rel}(a, G_i) + w_2 \text{fit}(a, d_i) + w_3 \text{var}(a, i, t) + w_4 \text{pref}(a, i) \quad (2)$$

The score has four terms. The term rel measures how directly the activity addresses the objective, fit measures how close its difficulty is to the learner's current level d_i , var rewards activities and artworks that have not been used recently, and pref reflects choices the learner has expressed. The weights are not learned. Educators choose among presets such as "practice the same skill" or "try something new", and the preset sets the weights. A first release will have too few users and too little data to learn weights reliably, and visible weights are easier

for teachers to trust. The top two or three activities are shown together with the terms that contributed most to their scores, and these terms form the explanation.

Difficulty is adapted with explicit rules instead of a learned policy. Let k be the number of consecutive successes with little prompting and m the number of consecutive difficulties, with thresholds κ and μ set by the educator:

$$d_i^{t+1} = \begin{cases} \min(d_i^t + 1, d_{max}) & \text{if } k \geq \kappa \\ \max(d_i^t - 1, d_{min}) & \text{if } m \geq \mu \\ d_i^t & \text{otherwise} \end{cases} \quad (3)$$

Difficulty here describes what the task demands and is not an estimate of the child's ability. A higher level means more choices, more language or more steps. Stepping down may also trigger an offer of a break or a change of format, and educators can freeze, reset or override d_i whenever they judge it appropriate.

Knowledge tracing models can estimate skill mastery from response sequences [32], and reinforcement learning can optimize instructional sequences [33]. Both need longitudinal data and carefully defined objectives that a first release will not have (Section 7.5). There is also an ethical difficulty. A reinforcement learning agent needs a reward signal, and defining one for a child with a communication disability is fraught: completion time penalizes slower responders, accuracy penalizes learners who explore, and engagement metrics can reward distraction as easily as learning. Rules are transparent and predictable. In our judgement that matters more than optimality when learners cannot easily report that the system is misjudging them.

4.3 Design propositions

Five propositions underlie the model. We state them so that future studies can test them, and the paper does not claim that they hold.

P1 predicts that offering more than one way to perceive and respond creates more opportunities for participation than a single-mode activity. The claim seems obvious, but it has not been tested with folk-art content and children with disabilities, so we state it as a hypothesis. P2 makes a similar prediction about context: showing verified provenance and reviewed cultural background next to an image should support better-informed discussion than showing the image alone.

P3 concerns agency. Letting learners choose, pause or change format should support their sense of control, although for some children with severe cognitive disabilities the range of meaningful choices may be narrow and the idea of autonomous selection becomes harder to apply.

P4 is about catching errors before they reach learners: review by educators and cultural practitioners should reduce unsupported or culturally inaccurate AI descriptions. P5 predicts

that visible reasons and alternatives make AI suggestions easier for educators to evaluate and override. Whether transparency changes how often educators actually override suggestions remains an empirical question.

4.4 Illustrative scenarios

Four hypothetical scenarios illustrate intended use. None of them reports an observed outcome.

In audio-supported exploration, a teacher selects a rights-cleared Đông Hồ print and presents the image with a short description and a reviewed audio description. The learner can replay the audio, point to a detail, or answer by speaking, choosing a symbol or drawing.

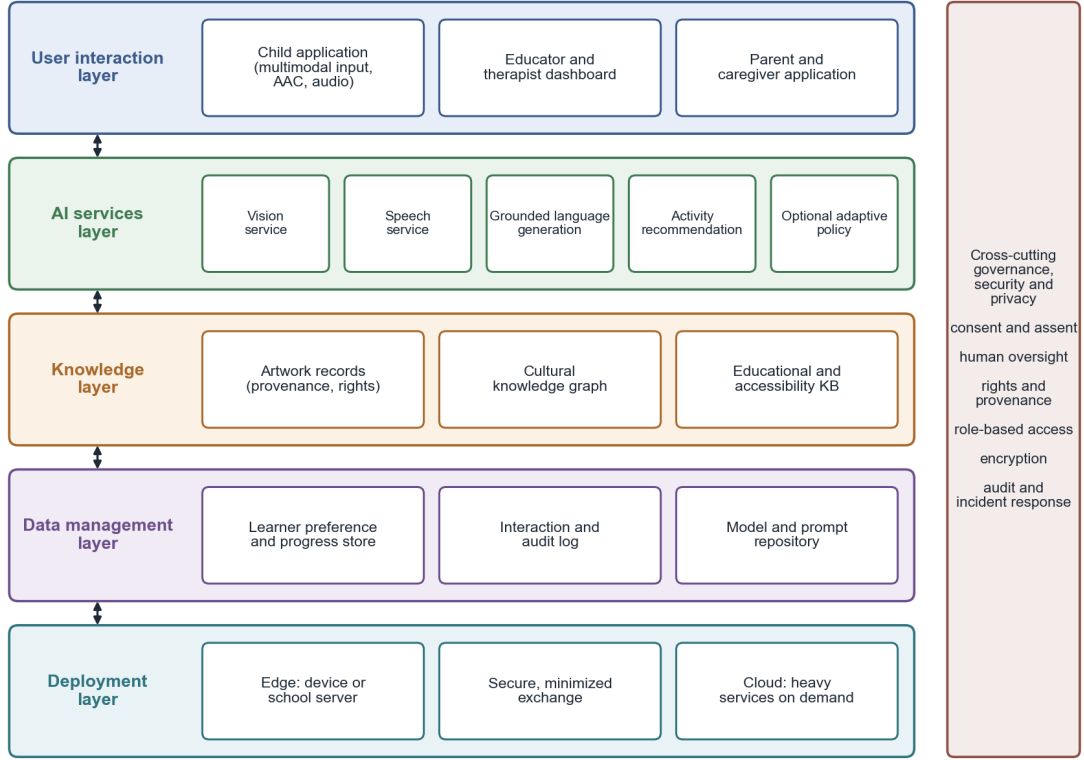
A tactile, object-based activity works differently. The teacher prints a raised-line outline or brings physical objects that match figures in the painting and explains that the adaptation shows only some features of the original. The platform provides preparation guidance, and the teacher decides whether the activity suits the learner.

In collaborative storytelling, a small group works around one painting and the children interact with each other as well as with the platform. Each child contributes through speech, drawing, picture cards, an AAC device or a partner's support, so no child is left in a passive role because of disability.

The fourth scenario is teacher planning. The teacher enters an objective, the access options needed, the time available and the materials at hand, and receives two or three activity variants, each with an explanation and a flag for any content still awaiting review.

5. Architecture overview

The architecture has five functional layers, stacked from the user interface down to deployment (Figure 2), and a governance column that cuts across all of them. Learners, educators and caregivers see only the top layer. Requests pass down to the AI services, which draw on the knowledge layer and reach data only through the data management layer, and the deployment layer decides where each component runs. Consent, oversight, rights management, access control, encryption and audit belong to the governance column.



Learners, educators, therapists and caregivers interact only through the top layer; every layer is subject to the governance column.

Figure 2. Layered architecture of the platform with a cross-cutting governance column.

A microservice architecture would make deployment more flexible, but service discovery, API versioning and network latency between components seem out of proportion for a platform that may serve a single school, so we use layers. Each layer boundary enforces a requirement. Separating AI services from the knowledge layer means that generated content can always be traced back to reviewed records (FR5, FR6). A dedicated data layer forces services to work with identifiers and parameters instead of whole learner profiles (NFR3), a constraint that would be very hard to add after the fact. Separating deployment lets sensitive processing stay on the device while heavier services move to the cloud only when needed (NFR2, NFR3). If one service fails or is switched off, the others keep running.

6. User interaction layer

6.1 Child application

Children explore paintings and complete activities in a child-facing application that supports touch, keyboard, switch access, symbol selection, drawing and speech, and an adult can operate the interface together with a child who needs physical or cognitive support. The interface is plain, with large targets, high contrast, adjustable text size and audio volume, predictable navigation and as little on screen as the activity allows. Nothing flashes. There is

no time pressure unless an educator sets one, there are no punitive scores or manipulative rewards, and a child who pauses, skips or repeats is never penalized.

A single painting can support several activity types. In object discovery, the learner finds a figure such as the buffalo or the rooster by touching it, choosing its name from a short list or saying it, and when the vision service is unsure the learner is offered choices instead of a judgement. Vocabulary activities pair an image region with a short label, optional audio and a plain-language description; the learner repeats the word, matches it to another image or signals understanding in another mode. Pronunciation feedback appears only when it matches the learner's communication goals and is never treated as a proxy for intelligence.

Sequencing divides a scene or a related set of prints into steps that the learner puts in order with picture cards or by dictating. Storytelling is more open: the learner narrates what is happening in a painting, and a language model may suggest a continuation. The learner's own interpretation is kept and never replaced by a single "correct" narrative, since the goal is expression. Fine-motor activities such as tracing or colouring a printed outline are offered only when they appear in an individual plan, with adjustable target sizes, alternative inputs and an easy way to stop. They do not replace occupational therapy.

Every activity template declares its objective, required modes, difficulty controls, access alternatives, stopping conditions and the data it records, and Equation (1) checks these declarations.

6.2 Educator and therapist dashboard

Educators plan, review and supervise through the dashboard. For each learner they set objectives, choose access settings and session limits, and review the platform's suggestions. A suggestion card shows the activity, the artwork, the reasons for its ranking and any content flagged as unreviewed. The educator can accept, edit or reject the suggestion, choose a different mode or replace it altogether.

Progress appears as a list of objectives, completed activities and observations recorded by adults. Where the dashboard summarizes change over time, it separates direct observations from AI-derived suggestions and labels the result as a descriptive record, not a clinical assessment. Draft reports for families or colleagues can be generated, but a professional must review each draft before it is shared, and no intervention is described as effective without evidence from an appropriate study. The dashboard does not compare a learner's performance with that of peers, because comparison data puts pressure on teachers to normalize children and conflicts with the principle of adapting tasks instead of labelling children.

AI-generated material also passes through the dashboard. New questions, simplified descriptions and story prompts wait in a review queue, nothing reaches a classroom without approval, and approved items are versioned so that a teacher can see which version a learner used.

6.3 Parent and caregiver application

A caregiver application lets learning continue at home. It offers optional activities linked to teacher-approved objectives, short summaries of school activities and practical ideas, such as retelling a print's story at dinner or finding similar animals around the house. It shares only information the school has authorized, and professional responsibility stays with the school.

7. AI services layer

Every AI service is modular and has a stated purpose, documented limits and an owner who reviews its behaviour. No service is required. When one is unavailable or switched off, static reviewed content takes its place. We treat this as a primary design requirement, since schools where internet access is unreliable must still be able to run lessons.

7.1 Vision service

The vision service takes digitized paintings and classifies them by tradition where this helps retrieval, detects and localizes figures and objects, extracts features for finding similar works and proposes candidate labels for human review.

The choice of model family depends on the images actually available, the rights attached to them and the computing budget; ResNet [23], Vision Transformers [24] and transformer-based detectors [25] are all candidates. Folk prints are stylized and look very different from the photographs most models are trained on, so accuracy cannot be taken for granted. Culturally specific symbols, such as a stylized buffalo with layered agricultural meaning, are harder still than the everyday objects these models were built to recognize. Detected objects therefore remain candidates until a human reviewer confirms them. Confirmed regions are stored in the artwork record and reused, so the vision model mostly runs when new artworks are added and rarely during lessons.

7.2 Speech service

The speech service has two functions. Text-to-speech reads descriptions, questions and instructions aloud in Vietnamese and, if needed, other languages; speech recognition converts spoken answers to text when a learner chooses to speak. Large multilingual recognizers trained on weakly supervised data [34] are a reasonable starting point, with an important caveat. Children's speech is generally harder for recognizers than adult speech, regional accents add to the difficulty, and atypical speech from children with motor or neurological differences raises error rates further. Accuracy for these groups must be verified before the function goes live.

Speech is never required. Every activity that accepts a spoken answer also accepts at least one other mode, since a system meant for children with motor and neurological differences cannot make voice input mandatory, however accurate the recognizer becomes. A recognition failure prompts another attempt or offers choices and never produces a wrong mark. Audio recordings stay on the device where possible and are not stored by default. This gives up

some analytical capability in exchange for a privacy guarantee that parents can understand without reading a technical document.

7.3 Grounded language generation

A language model drafts questions at different levels, simplified descriptions, alternative text, activity instructions and story prompts. Transformer-based language models can produce such text from a few examples [35, 36], and they also hallucinate, producing fluent statements that are false [6].

We use retrieval-augmented generation [37] instead of fine-tuning. A fine-tuned model might write more fluent Vietnamese, but every update could change how it describes a culturally sensitive image, and a generated description could not be traced back to a specific source. Retrieval-augmented generation gives up some fluency for auditability. For each request, the service retrieves the relevant artwork record and linked context from the knowledge layer and instructs the model to use only that material. The inputs are approved artwork metadata, verified context, the learning objective and the access settings, and the outputs are drafts that cite the records they drew on.

Three safeguards apply. Generated text is always marked as generated and never mixed with verified cultural information. Drafts enter the educator review queue before classroom use, so nothing reaches a child unreviewed. When retrieval finds no reviewed source for a requested fact, the service does not invent context; it falls back to a short verified description followed by an open question such as "What do you see the animals doing?", which still gives the learner something to respond to. Prompt templates and the model version used for each draft are logged in the repository (Section 9.3), and no information about a learner's diagnosis or other sensitive attributes is ever passed to this service.

7.4 Activity recommendation service

Equations (1) and (2) are implemented as a knowledge-based recommender in Burke's taxonomy [38], which reasons over declared properties of activities and stated constraints instead of learning from the behaviour of many users. With few users per school, sensitive data and a strong need for explanation, this is the appropriate approach. Candidates include visual exploration and object matching, audio-supported storytelling, tactile or printable activities, symbol-supported communication tasks and collaborative art and narrative activities, each at several levels of interaction complexity.

Explanations come directly from the ranking terms, for example: "Suggested because animal vocabulary is a current goal, the previous matching activity was completed with little prompting, and this print has not been used this week." A teacher who disagrees can point to the exact term that produced a suggestion they would not have made. If a learned ranking model later replaced the rules, SHAP could produce comparable explanations [39], but the current design does not need it.

Interaction time is a poor measure of benefit. Long interaction can mean confusion as easily as engagement, and short interaction can mean the interface was inaccessible. The platform therefore does not use either as a direct measure of benefit.

7.5 Optional adaptive policy

Reinforcement learning could in principle replace rule-based adaptation with a learned sequencing policy [33], where the state describes the current activity and learner context, actions include format changes or difficulty adjustments, and the reward encodes educational objectives. Whether this would help is unclear. A review of reinforcement learning for instructional sequencing found that results depend heavily on how reward and student models are defined, and many studies did not outperform simpler baselines [40].

The component is therefore switched off. Enabling it requires consented longitudinal data, a reward definition reviewed by educators and therapists, offline evaluation against the rule-based policy of Equation (3) and a check for unequal effects across groups of learners. The bar is high on purpose. Two rules apply without exception: the policy never optimizes for compliance or time on task, and it always yields to human override.

7.6 Affect-related support

The platform does not infer emotions from faces, voices or behaviour. Barrett et al.'s review shows that the evidence for reading emotions from facial movements is weak in general [5], and for children with disabilities, whose expressions, gaze and movements may differ from those in training populations, the risk of misreading outweighs any benefit from detecting frustration or disengagement automatically.

The design relies on direct signals instead. A learner can press a "break" or "too hard" button or choose a feelings symbol, or an adult can record an observation, and any of these can trigger a pause, a simpler format or a change of activity. If an affective module were ever added, it would need its own ethical and technical assessment, explicit consent, independent validation and a rule barring its outputs from diagnosis, grading or discipline.

7.7 Multimodal integration and graceful degradation

Services are combined by late fusion: each produces its own output and the activity logic decides how to use it, instead of a single model receiving raw features from all modalities as in early fusion [41]. Late fusion has two advantages that matter here more than raw performance. Any modality can be missing or disabled without breaking a session, which is a normal situation in a classroom of non-speaking children where the speech service may be switched off altogether, and each output can be checked and explained on its own.

Each service has a fallback. Without speech recognition, learners use touch or symbols. Without language generation, reviewed descriptions are shown. Without vision, the platform uses regions already confirmed in the artwork record. Without a network, the device runs cached activities, which is how the platform works in schools with unreliable connections.

Because degradation is part of the core design, it does not depend on resilience features that could be dropped under schedule pressure.

8. Knowledge layer

8.1 Artwork record schema

At the core of the knowledge layer is one record per artwork, whose minimum fields are listed in Table 3. Verified metadata are kept separate from proposed annotations, and each annotation records who created it and in which role, along with its review status and version. Provenance fields should point to sources such as Durand's documentation of the traditions [8] and institutional records. Following the datasheet model [42], the composition and limits of any image collection should also be documented, so that later users can see what is and is not represented.

Table 3. Minimum fields of an artwork record.

Field group	Fields	Notes
Identity	artwork_id, title, title_variants, tradition, region	Title variants keep Vietnamese names with diacritics
Provenance and rights	source_institution, source_reference, provenance, rights_status, permitted_uses	Released only after rights are confirmed
Image	image_uri, image_quality, digitization_method	Authorized reproduction only
Visual annotations	object_regions, object_labels, colours, composition_notes	Candidates from vision service until confirmed
Cultural context	cultural_description, linked_context_ids, interpretation_attribution	Interpretations always carry a source
Educational metadata	vocabulary, learning_objectives, suitable_activity_types	Proposed by educators
Accessibility	alt_text, audio_description, visual_complexity, plain_language_summary	Reviewed before use
Review	annotation_role, review_status, reviewer_role, version	Records who proposed and who approved

Some fields need care. "Festive scene" describes content and says nothing about how a child viewing it feels. Tags for therapeutic use indicate activity categories a professional might choose and promise no benefit. Sinh prints are often devotional or ritual, so they carry a flag that requires cultural review before any classroom use; a worship image used in a vocabulary exercise could trivialize its meaning in ways that someone outside the tradition might not notice. Child data never enters the knowledge layer. Speech, video and interaction logs belong to the data management layer, where access controls and retention limits apply.

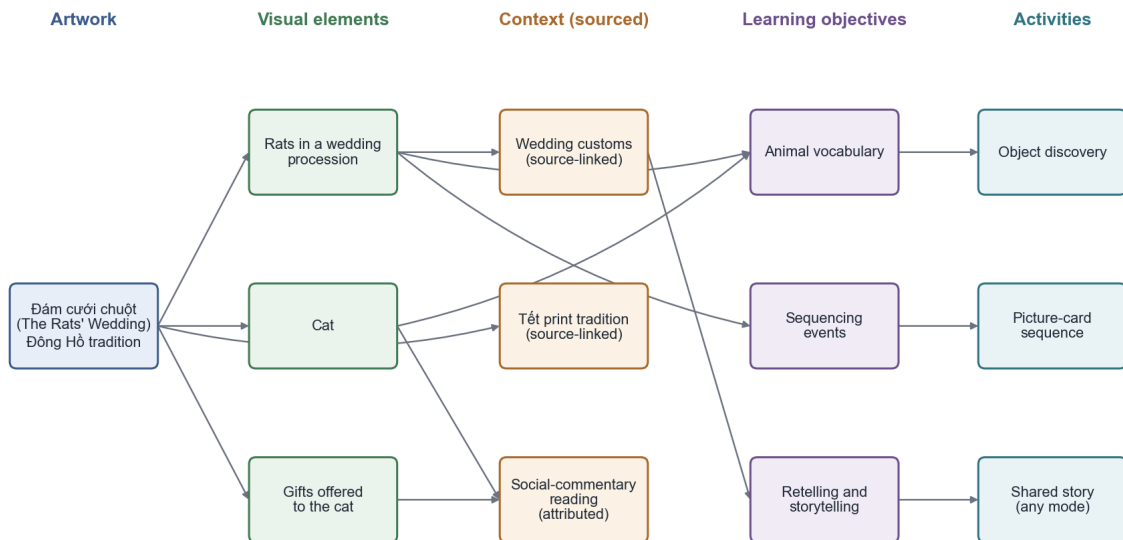
Đám cưới chuột (The Rats' Wedding) from the Đông Hồ tradition shows how this works. Candidate annotations would mark the rats in the procession, the cat and the gifts the rats offer. A separate context record could describe readings of the scene's social commentary, provided each reading is supported by a credible source and attributed to it. Observable content and interpretation stay distinct.

8.2 Cultural knowledge graph

A relational database could store the same information and might be simpler to maintain for a small collection. We use a graph because the queries the platform needs, such as "find all artworks linked to a learning objective through a cultural context that has been reviewed", are path queries, and in SQL they require joins that obscure their intent. The cultural knowledge graph connects artworks, visual elements, sourced context, learning objectives and activities through typed nodes and edges [27] and borrows from heritage ontologies the practice of recording provenance and events explicitly [26]. A typical path runs:

Artwork → Visual element → Sourced cultural context → Learning objective → Accessible activity

Figure 3 shows a fragment for Đám cưới chuột. Every edge stores its source and a confidence label, and interpretive nodes are marked as attributed, so several readings of the same image can coexist. When the language service retrieves context, it follows only edges whose sources have been reviewed. A documented depiction of a buffalo in a particular print, for example, can be linked to agricultural life if a reliable source supports the link, and that link can then lead to a vocabulary or storytelling activity about traditional livelihoods. Where meanings differ between regions or sources, the graph keeps the variation.



Each edge stores its source and a confidence label; interpretive nodes are marked as attributed and can coexist with alternative readings.

Figure 3. Fragment of the cultural knowledge graph for Đám cưới chuột (The Rats' Wedding). Observable elements, sourced context, learning objectives and activities are separate node types.

8.3 Educational and accessibility knowledge base

A third store holds reviewed activity templates, UDL-aligned teaching strategies, AAC support options, accessibility requirements for digital and printed materials, safety considerations and criteria for when human review is needed. Its content is written or approved by qualified educators and therapists. General educational guidance belongs here and clinical protocols do not; those remain the responsibility of professionals and their institutions, and encoding them as automatic rules would suggest a clinical authority that the platform does not have.

9. Data management and deployment

Because the users are children, every decision in the data layer follows from four rules: minimize data, limit purposes, control access and keep records only as long as necessary.

9.1 Learner preference and progress store

The information held about learner i at time t is a small tuple:

$$P_i(t) = \left(M_i, G_i(t), \left\{ d_{i,g}^t \right\}_{g \in G_i(t)}, H_i^\tau(t), N_i(t) \right) \quad (4)$$

Here M_i holds the modes and presentation formats the learner has selected, $G_i(t)$ the objectives set by educators, $d_{i,g}$ the current task difficulty for each objective, $H_i^\tau(t)$ the activity history within a retention window τ , and $N_i(t)$ notes authorized by the responsible professional. The profile contains no diagnosis, no biometric data and no facial images. Preferences are recorded at the level of activities and sessions and are never turned into permanent assumptions about the child. Educators and, where appropriate, families can view, correct and delete entries at any time.

9.2 Interaction and audit log

The log records the activities offered and chosen, format changes made by users, recommendations with their explanations, educators' approvals, edits and rejections, and technical or accessibility errors. Emotional states are not logged by default. Entries carry pseudonymous identifiers, follow retention limits and are visible only to authorized roles. The log exists for accountability and is not used to monitor children's behaviour.

9.3 Model and prompt repository

A repository stores model identifiers, versions, training metadata, prompt templates, evaluation protocols, documented limitations, validation status and change history. Each model has a card stating its intended use and known limits [43]. MLflow or a similar registry can manage these records. Before any model or prompt version goes live, testing must cover accessibility, cultural accuracy, privacy and possible harm, since technical metrics alone are not enough, and every release can be rolled back.

9.4 Cloud-edge deployment

Each function runs in the simplest environment that can handle it safely (Figure 4). On the edge, meaning a learner's device or a local school server, the platform renders the interface and accessibility controls, plays reviewed audio, keeps the activity state and session timer, caches approved content and rules, and runs lightweight speech recognition where the hardware allows. Sensitive audio and images leave the edge only when sending them is necessary, authorized and protected.

Heavier services run in the cloud: grounded language generation, knowledge services, the model and content registry, aggregated monitoring and the review queue for generated content. Each operation sends only what it needs, usually artwork identifiers and activity parameters. Raw child audio, facial images and detailed histories stay on the edge. Any future model training would follow its own governance rules on data rights, consent, security and deletion, separate from ordinary operation.

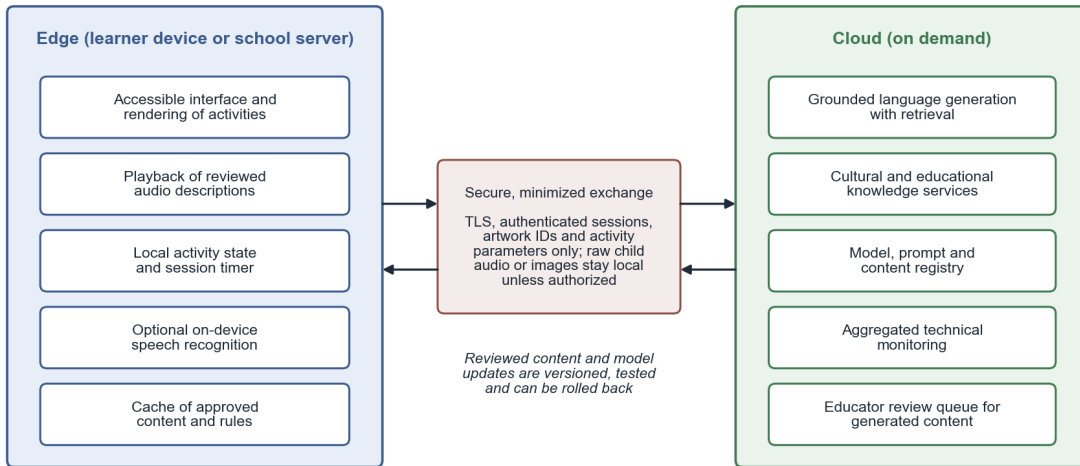


Figure 4. Cloud-edge deployment. Sensitive inputs stay on the edge; the cloud handles heavier services on demand, and reviewed updates are synchronized back in versioned form.

Communication between edge and cloud is encrypted and authenticated, and every request passes an authorization check. Updates to content and models are versioned, tested and reversible and are synchronized to the edge under controlled conditions. Many interactions need no cloud at all, and a session using cached activities works offline. Offline operation is treated as a full mode of the platform, because a design that assumes constant cloud access would fail in schools where connectivity is unreliable.

10. Governance, privacy, security and ethics

10.1 Consent, assent and participation

Under the CRC and the CRPD, children with disabilities are entitled to extra protection and to a voice in decisions that affect them [2, 14]. Parental permission alone is therefore not

enough. The child's own assent should be sought in a form the child can understand and express, whether through speech, symbols or AAC; how to obtain meaningful assent from a child with severe cognitive disability is a problem this design does not solve. Refusing or stopping has no cost for the child, who keeps access to other activities and services, and families and professionals decide together which functions are switched on.

10.2 Data minimization and pseudonymization

The platform collects as little as possible. Facial video, voice recordings, eye-gaze data and detailed behavioural traces enter the system only when a specific purpose requires them and that purpose has been documented in advance. Names and contact details are kept in a separate store from educational records and linked through pseudonymous identifiers, whose keys are held under restricted access by a designated data controller. If computer vision ever has to process an image of a child, which the core design avoids, task-specific outputs are kept instead of raw images wherever possible.

Pseudonymization does not make data anonymous. Images, voice recordings and derived features can still identify a person or reveal sensitive information, so retention limits, restrictions on secondary use and secure deletion all apply. In Viet Nam the Law on Personal Data Protection has been in force since 1 January 2026 [44], and the GDPR applies where European data subjects are involved [45]. Identifiable child data is never released publicly.

10.3 Secure transmission and access control

Data in transit is encrypted, and sensitive data at rest is encrypted on devices and servers. Authentication differs by user group: children use a simple supervised sign-in, while adults who access records need stronger authentication. Role-based access control gives each role only what it needs, so a teacher sees progress for their own learners, while a technical administrator who maintains the server does not automatically see any educational records. Secure session management, audit logging, timely updates, vulnerability assessment and an incident response plan complete the picture. All of this has to fit how schools actually work, with shared devices and limited technical staff.

10.4 Privacy-preserving learning options

If the platform later improves its models with data from several schools, privacy-preserving techniques can reduce exposure. In federated learning, each site trains locally and a server aggregates the updates [46]:

$$w^{t+1} = \sum_{k=1}^K \frac{n_k}{n} w_k^{t+1} \quad (5)$$

where w_k^{t+1} is the model trained at site k on n_k local examples and n is the total. Model updates can still leak information, so federated learning can be combined with secure aggregation, which lets the server compute the sum without seeing individual updates [47], and with differential privacy, which bounds how much any one person's data can change the

output. A mechanism $\mathcal{M}$ is ϵ -differentially private if, for any two datasets D and D' that differ in one person and any set of outputs S [48]:

$$Pr[\mathcal{M}(D) \in S] \leq e^\epsilon Pr[\mathcal{M}(D') \in S] \quad (6)$$

The data owners have to choose ϵ , trading protection against usefulness. These techniques complement each other and suit different threat models. None of them is part of the first release; they are options for a later stage, when sharing data across schools becomes practical.

10.5 Cultural governance and rights

Provenance, reproduction rights, title variants, regional meanings and restrictions on ritual images all require consultation with cultural institutions, artisans and community representatives. Digitizing an image transfers neither copyright nor cultural authority. Sources must be credited, permitted uses respected, and cultural reviewers must be able to correct or withdraw content. Only a few families still practise Đông Hồ printmaking [4], and reproducing their work digitally without involving them would be inappropriate whatever its copyright status. Sinh prints are made for ritual purposes [10], so displaying them in a classroom activity risks trivializing their meaning, and consultation on their use has to be substantive.

10.6 Risk register

Table 4 lists eight risks with their mitigations and the role responsible for each. The register is reviewed whenever a service, model or content set changes. Its common thread is that the platform supports teachers, therapists, families and clinical care and never replaces them [20, 49].

Table 4. Risk register.

Risk	Mitigation	Owner
Invented cultural explanations	Retrieval from reviewed records only; review queue; attribution	Content lead with cultural reviewer
Incorrect object labels	Human confirmation of candidates; choices offered when uncertain	Content lead
Speech recognition failure	Speech never required; alternative modes; no penalty	Accessibility lead
Inaccessible activities	WCAG 2.2 audit; testing with assistive technologies	Accessibility lead
Adult over-reliance on suggestions	Explanations, alternatives and reminders that decisions belong to educators	Pedagogical lead
Privacy breach	Minimization, encryption, role-based access, incident plan	Data protection officer
Unapproved artwork use	Rights fields checked before release; withdrawal process	Content lead
Profile bias from short-term data	Task-specific, reversible profiles; reset and correction	Pedagogical lead

11. Discussion

11.1 Contribution

The individual components are familiar. UDL is well established, the CIDOC ontology has been used in the heritage sector for two decades, and speech recognition and language generation are widely available. What this paper adds is a way of combining them: a buildable design that brings cultural heritage, accessibility and AI into one platform where adults stay in control. Reviewed artworks become activities that can be reached through several modes, generated text is tied to reviewed records, and every suggestion passes an educator before it reaches a child. Since the selection and adaptation rules are explicit, the platform's behaviour can be predicted and explained. Separating observation from interpretation in the knowledge graph gives cultural reviewers a clear boundary to work at, which matters most when the same image carries different meanings in different regions.

11.2 Tensions and trade-offs

Several tensions shaped the design, and in each case we chose caution over capability.

Cameras and continuous audio would yield richer interaction signals, but the privacy cost is too high for children who cannot meaningfully consent, so the design leaves them out. A learned adaptation policy might save teachers setup time; rules they can read and override are easier to trust in a first release, and that release will largely decide whether teachers adopt the platform. Generative storytelling invites expression and can also distort cultural meaning, which is why generation stays grounded in reviewed records and passes through an educator.

Running models on the device protects data and works offline but limits model size, so heavier functions move to the cloud only when needed.

The platform does not generate new images in the style of folk paintings. Generated images have no provenance, and imitating the style of a living tradition without the involvement of its artisans would be inappropriate.

Traditional painting is not inherently therapeutic. Art-based activities may be enjoyable and useful within a broader plan, but whether they benefit a child depends on the child, the activity, how it is delivered and what is measured, and engagement alone is not evidence of progress.

11.3 Limitations

Nothing in this paper has been validated. We do not know whether children will prefer folk-painting activities to more familiar digital formats, whether vision models will cope well enough with stylized prints to be useful, or whether the platform will improve learning in any measurable way. We expect cultural familiarity to make engagement easier, but that is a hypothesis to be tested.

Cultural interpretation is the hardest part. It varies by region, community, version of a print and historical period, and a single annotation scheme risks flattening layered meanings. Community review addresses this but is slow and expensive. AI components will perform unevenly across speech patterns, access technologies and individual children. Short-term performance, if treated as a stable trait, biases the learner profile; task-specific and correctable records counter this, though perhaps not completely. A privacy-preserving implementation adds cost. And the rights situation for many historical prints is unclear, which could shrink the usable collection considerably.

12. Conclusion and future work

We have described one way to connect Vietnamese folk paintings to inclusive learning through AI, with adults in control and children's data kept to a minimum. The architecture spans curated artwork records, a provenance-aware knowledge graph, multimodal AI services and a cloud-edge deployment, with governance built into every layer. Whether it works for children with disabilities, their teachers and cultural custodians remains open, and only implementation and evaluation can answer that.

The work should proceed in stages. Artwork rights come first: they have to be established with institutions and a small reviewed collection assembled, which may take months given how fragmented ownership of historical prints is. Activities can then be co-designed with educators, families and learners, and a rule-based prototype built. Usability studies under ethical review should come before any controlled evaluation of outcomes. AI services should be introduced one at a time, with language generation first, speech recognition next and adaptive sequencing last, each with safety checks and a fallback ready. Throughout, the aim is

to make culturally grounded material more accessible without taking judgement away from the people who know the child.

References

- [1] UNESCO. Global Education Monitoring Report 2020: Inclusion and education: All means all. Paris: UNESCO; 2020.
- [2] United Nations. Convention on the Rights of Persons with Disabilities. New York: United Nations; 2006.
- [3] World Health Organization, United Nations Children's Fund (UNICEF). Global report on assistive technology. Geneva: WHO and UNICEF; 2022.
- [4] UNESCO. Craft of making Đông Hồ folk woodblock printings. List of Intangible Cultural Heritage in Need of Urgent Safeguarding, Decision 20.COM 7.a.1; 2025. Available from: <https://ich.unesco.org/en/USL/01737>
- [5] Barrett LF, Adolphs R, Marsella S, Martinez AM, Pollak SD. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychol Sci Public Interest*. 2019;20(1):1–68. doi:10.1177/1529100619832930
- [6] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv*. 2023;55(12):248. doi:10.1145/3571730
- [7] Holmes W, Porayska-Pomsta K, Holstein K, Sutherland E, Baker T, Buckingham Shum S, et al. Ethics of AI in education: Towards a community-wide framework. *Int J Artif Intell Educ*. 2022;32(3):504–526. doi:10.1007/s40593-021-00239-1
- [8] Durand M. Imagerie populaire vietnamienne. Paris: École française d'Extrême-Orient; 1960. (Publications de l'École française d'Extrême-Orient, vol. 47).
- [9] Viet Nam News. Traditional folk painting enjoys a revival. Viet Nam News; 13 February 2019. Available from: <https://vietnamnews.vn/life-style/505141/traditional-folk-painting-enjoys-a-revival.html>
- [10] Viet Nam News. Huế culture shines in Sinh folk paintings. Viet Nam News; 8 February 2025. Available from: <https://vietnamnews.vn/sunday/features/1691764/hue-culture-shines-in-sinh-folk-paintings.html>
- [11] CAST. Universal Design for Learning Guidelines version 3.0. Lynnfield, MA: CAST; 2024. Available from: <https://udlguidelines.cast.org>
- [12] World Wide Web Consortium (W3C). Web Content Accessibility Guidelines (WCAG) 2.2. W3C Recommendation; 5 October 2023.
- [13] Light J, McNaughton D. Communicative competence for individuals who require augmentative and alternative communication: A new definition for a new era of communication? *Augment Altern Commun*. 2014;30(1):1–18. doi:10.3109/07434618.2014.885080
- [14] United Nations. Convention on the Rights of the Child. New York: United Nations; 1989.
- [15] National Assembly of Viet Nam. Law on Persons with Disabilities No. 51/2010/QH12. 2010.
- [16] National Assembly of Viet Nam. Law on Children No. 102/2016/QH13. 2016.
- [17] Hopcan S, Polat E, Ozturk ME, Ozturk L. Artificial intelligence in special education: A systematic review. *Interact Learn Environ*. 2023;31(10):7335–7353. doi:10.1080/10494820.2022.2067186
- [18] Zdravkova K, Krasniqi V, Dalipi F, Ferati M. Cutting-edge communication and learning assistive technologies for disabled children: An artificial intelligence perspective. *Front Artif Intell*. 2022;5:970430. doi:10.3389/frai.2022.970430
- [19] Kasneci E, Seßler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learn Individ Differ*. 2023;103:102274. doi:10.1016/j.lindif.2023.102274
- [20] UNICEF. Policy guidance on AI for children, version 2.0. New York: UNICEF; 2021.
- [21] Fiorucci M, Khoroshiltseva M, Pontil M, Traviglia A, Del Bue A, James S. Machine learning for cultural heritage: A survey. *Pattern Recognit Lett*. 2020;133:102–108. doi:10.1016/j.patrec.2020.02.017
- [22] Cetinic E, She J. Understanding and creating art with AI: Review and outlook. *ACM Trans Multimedia Comput Commun Appl*. 2022;18(2):66. doi:10.1145/3475799
- [23] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016. p. 770–778. doi:10.1109/CVPR.2016.90
- [24] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: *International Conference on Learning Representations (ICLR)*; 2021.

- [25] Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A, Zagoruyko S. End-to-end object detection with transformers. In: *Computer Vision – ECCV 2020. Lecture Notes in Computer Science*, vol 12346. Cham: Springer; 2020. p. 213–229. doi:10.1007/978-3-030-58452-8_13
- [26] Doerr M. The CIDOC conceptual reference module: An ontological approach to semantic interoperability of metadata. *AI Mag.* 2003;24(3):75–92.
- [27] Hogan A, Blomqvist E, Cochez M, d'Amato C, de Melo G, Gutierrez C, et al. Knowledge graphs. *ACM Comput Surv.* 2021;54(4):71. doi:10.1145/3447772
- [28] Hevner AR, March ST, Park J, Ram S. Design science in information systems research. *MIS Q.* 2004;28(1):75–105. doi:10.2307/25148625
- [29] Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. *J Manag Inf Syst.* 2007;24(3):45–77. doi:10.2753/MIS0742-1222240302
- [30] Shneiderman B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int J Hum Comput Interact.* 2020;36(6):495–504. doi:10.1080/10447318.2020.1741118
- [31] Amershi S, Weld D, Vorvoreanu M, Fournery A, Nushi B, Collisson P, et al. Guidelines for human-AI interaction. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019. Paper 3. doi:10.1145/3290605.3300233
- [32] Corbett AT, Anderson JR. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model User-Adapt Interact.* 1994;4(4):253–278. doi:10.1007/BF01099821
- [33] Sutton RS, Barto AG. Reinforcement learning: An introduction. 2nd ed. Cambridge, MA: MIT Press; 2018.
- [34] Radford A, Kim JW, Xu T, Brockman G, McLeavey C, Sutskever I. Robust speech recognition via large-scale weak supervision. In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR 202; 2023. p. 28492–28518.
- [35] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems* 30. 2017. p. 5998–6008.
- [36] Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. In: *Advances in Neural Information Processing Systems* 33. 2020. p. 1877–1901.
- [37] Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In: *Advances in Neural Information Processing Systems* 33. 2020. p. 9459–9474.
- [38] Burke R. Hybrid recommender systems: Survey and experiments. *User Model User-Adapt Interact.* 2002;12(4):331–370. doi:10.1023/A:1021240730564
- [39] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems* 30. 2017. p. 4765–4774.
- [40] Doroudi S, Aleven V, Brunskill E. Where's the reward? A review of reinforcement learning for instructional sequencing. *Int J Artif Intell Educ.* 2019;29(4):568–620. doi:10.1007/s40593-019-00187-x
- [41] Baltrušaitis T, Ahuja C, Morency L-P. Multimodal machine learning: A survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell.* 2019;41(2):423–443. doi:10.1109/TPAMI.2018.2798607
- [42] Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé III H, et al. Datasheets for datasets. *Commun ACM.* 2021;64(12):86–92. doi:10.1145/3458723
- [43] Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. 2019. p. 220–229. doi:10.1145/3287560.3287596
- [44] National Assembly of Viet Nam. Law on Personal Data Protection No. 91/2025/QH15. Adopted 26 June 2025; effective 1 January 2026.
- [45] European Parliament and Council of the European Union. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation). *Off J Eur Union.* 2016;L119:1–88.
- [46] McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*. PMLR 54; 2017. p. 1273–1282.
- [47] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 2017. p. 1175–1191. doi:10.1145/3133956.3133982
- [48] Dwork C, Roth A. The algorithmic foundations of differential privacy. *Found Trends Theor Comput Sci.* 2014;9(3–4):211–407. doi:10.1561/04000000042
- [49] UNESCO. Recommendation on the ethics of artificial intelligence. Paris: UNESCO; 2021.