

**AR Educational Suite for Primary Students’
Geometry
and Literacy.**

R26-IM-005

Project Proposal Report

Wickramasingha N.N. – IT22214102

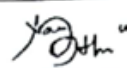
B.Sc. (Hons) Degree in Information Technology Specialized in
Interactive Media

Department of Information Technology
Sri Lanka Institute of Information Technology Sri Lanka

March 2026

DECLARATION

I declare that this is my own work, and this proposal does not knowingly incorporate any material previously submitted for a degree or diploma at any other university or higher education institution, nor does it contain any material that has been previously published or written by another person apart from where proper acknowledgment is given in the text.

Name	Student ID	Signature
Wickramasingha N.N	IT22214102	

The above candidate is conducting research for their undergraduate Dissertation under my supervision.

Name	Date	Signature
Mr. Aruna Ishara Gamage	10/03/2016	

ABSTRACT

Early literacy learning plays a vital role in children’s cognitive and academic development, yet many primary school students struggle to connect visual recognition, word formation, and pronunciation within a single learning process. Most existing educational tools focus only on isolated activities such as handwriting practice, pronunciation exercises, or object visualization, and therefore do not provide a unified multisensory learning experience. This research proposes a Multisensory Phonics Integration component within a mobile-based Artificial Intelligence (AI) and Augmented Reality (AR) educational system for primary school students aged 6–12. The proposed component is designed to combine student-generated sketch recognition, sketch-based 3D transformation, word matching, pronunciation validation, and AR-based reward feedback into one interactive workflow. In this process, the learner first draws a familiar object as a 2D sketch, after which the system recognizes the drawing and transforms it into a 3D representation based on the learner’s own sketch. The learner then writes the English name of the object using air-writing and pronounces the same word aloud. When the sketch, word, and pronunciation are correctly matched, the system generates a completed 3D model as an interactive AR reward using Unity and ARCore. By combining computer vision, sketch understanding, speech recognition, and AR rendering technologies, the proposed solution aims to create a low-cost, engaging, and feedback-driven literacy learning environment that improves vocabulary association, pronunciation practice, and learner motivation while also demonstrating potential for use in both classroom and home-learning settings.

Keywords: *Augmented Reality, Multisensory Learning, Phonics Learning, Sketch Recognition, Speech Recognition*

ACKNOWLEDGEMENT

I would like to express my sincere appreciation to all those who have contributed to the successful completion of this research proposal.

First and foremost, I would like to convey my deepest gratitude to my supervisor, and my co-supervisor, for their invaluable guidance, insightful suggestions, and continuous support throughout the preparation of this research proposal. Their constructive feedback, encouragement, and academic direction greatly contributed to refining the Multisensory Phonics Integration module and improving the overall quality of this work.

I would also like to extend my gratitude to the academic staff of the Faculty of Computing, Sri Lanka Institute of Information Technology (SLIIT), for providing the knowledge, resources, and academic foundation that enabled me to undertake this research. The skills and understanding gained during my studies played a significant role in shaping this proposed research on an AR Educational Suite utilizing Sketch-to-Object recognition and audio-visual phonics feedback.

My sincere thanks are also extended to my research group members for their cooperation, valuable discussions, and support in developing the overall research concept and system architecture. Their collaboration helped strengthen the integrated cross-domain loop, linking geometry validation and air-writing trajectory tracking with my phonics integration module.

Finally, I would like to express my heartfelt gratitude to my parents, family, and friends for their continuous encouragement, patience, and understanding throughout this research journey. Their support and motivation have been a constant source of strength in completing this proposal.

TABLE OF CONTENT

Table of Contents

DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENT	v
LIST OF TABLES	vi
LIST OF FIGURES.....	vi
LIST OF ABBREVIATIONS	vii
LIST OF APPENDICES	viii
1. INTRODUCTION.....	1
1.2 Literature Review	3
2. OBJECTIVE	10
2.1 Main Objective	10
2.2 Specific Objectives	11
3. METHODOLOGY	12
4. HIGH-LEVEL SYSTEM ARCHITECTURE	20
5. USER REQUIREMENTS	27
6. COMMERCIALIZATION PLAN.....	32
6.2 Value Proposition.....	33
6.3 Revenue Model.....	34
6.4 Cost Estimation	35
6.5 Pricing Strategy.....	35
6.6 Competitive Advantage	35
6.7 Intellectual Property Considerations	36
7. BUDGET AND JUSTIFICATION	37
8. WORK BREAKDOWN STRUCTURE (WBS)	39
9. GANTT CHART	40
10. REFERENCES LIST.....	41

11. APPENDICES	43
-----------------------------	-----------

LIST OF TABLES

Table 1: Specific Objectives of the Research.....	11
Table 3: Target Market Segments	32
Table 4: Customer Personas	33
Table 5:Cost Category.....	35
Table 6: Estimated Budget for the Research Project	37
Table 7: Budget Justification for the Research Project	38
Table 8: Work Breakdown Structure for Multisensory Phonics Integration Component	39
Table 9:9.GANTT CHART	40
Table 10: Observation Sheet	43
Table 11:Student Feedback Questions	44
Table 12:Preliminary UI Mockups.....	45
Table 13: Risk Analysis	46

LIST OF FIGURES

Figure 1:HIGH-LEVEL SYSTEM ARCHITECTURE	20
Figure 2 : System	22
Figure 3:Individual Component and Its Parts	23

LIST OF ABBREVIATIONS

AR Augmented Reality

CNN Convolutional Neural Network

RNN Recurrent Neural Network

ML Machine Learning

CV Computer Vision

MR Mixed Reality

HCI Human Computer Interaction

UI User Interface

UX User Experience

LIST OF APPENDICES

Appendix A – Survey Instruments

Appendix B – Detailed Diagrams

Appendix C – Preliminary UI Mockups

Appendix D – Risk Analysis

Appendix E – Ethical Documents

1. INTRODUCTION

Literacy education is one of the most important foundations of a child's academic development. Skills such as phonics recognition, handwriting accuracy, spelling, and vocabulary acquisition play a major role in helping children develop reading and writing proficiency. When these basic literacy skills are not properly developed during the early stages of schooling, students may experience long-term difficulties in language learning, reduced motivation, and poor performance in subjects that depend on reading comprehension and written communication.

One of the major challenges in primary education is that young learners often struggle to connect what they see, what they write, and what they say during the learning process. Phonics learning requires children to understand the relationship between letters, sounds, and words, while handwriting development requires practice in letter formation and spelling. In many classrooms, teachers must manage a large number of students at once, making it difficult to provide individual feedback on each child's writing, spelling, and pronunciation. As a result, students may develop weak phonics understanding, incorrect spelling habits, poor pronunciation, and limited word-object association.

This issue is especially significant among primary school learners aged 6 to 12 years, because children at this stage learn more effectively through interactive, visual, and activity-based learning methods than through passive instruction. However, traditional teaching methods still rely mainly on textbooks, blackboard explanations, worksheets, and repetitive writing exercises. Although these methods provide basic instructional support, they often do not create an engaging environment that allows children to actively participate, self-correct, and receive immediate feedback.

Recent developments in digital technology have created new possibilities for improving early literacy learning. In particular, Augmented Reality (AR) has emerged as an effective educational technology because it can overlay digital content onto the real world and create interactive learning experiences. Research in AR-based education has shown that AR can improve learner motivation, engagement, and conceptual understanding by allowing students to interact with digital objects in real time [1], [2].

These benefits are especially important for young learners, who usually respond positively to visual and game-like educational experiences.

However, existing AR educational systems are mostly designed for content visualization rather than for interactive literacy skill development. Many AR applications simply display 3D objects or animations without analyzing the learner's own input or validating learning performance [1], [2]. Similarly, many literacy learning applications focus only on one activity at a time, such as tracing letters, practicing spelling, or listening to word pronunciation. These systems usually do not combine multiple learning modes into a single integrated workflow.

Another limitation can be seen in current handwriting and air-writing research. Studies in trajectory-based handwriting recognition have shown that writing should not be evaluated only by the final visual result, but also by the movement path and sequence used during writing [5], [6]. In addition, advances in hand gesture recognition and pose estimation have demonstrated that computer vision can accurately detect finger and hand movements in real time [7], [8], [13], [14]. These developments provide a strong technical foundation for interactive literacy applications, but they are often applied in general gesture recognition or human-computer interaction settings rather than in child-centered educational systems.

At the same time, speech recognition technologies have significantly improved with the support of deep learning methods. Research has shown that recurrent neural network-based speech models can recognize spoken words accurately and can be used in educational applications such as pronunciation evaluation and language learning support [9]. Similarly, deep learning models for image and object recognition have achieved strong performance in recognizing visual patterns and shapes, making them useful for sketch interpretation and object understanding [10], [11], [12], [15]. These advances suggest that computer vision, speech recognition, and AR can be meaningfully combined to support literacy learning in a more interactive and intelligent way.

Based on these developments, this research proposes a Multisensory Phonics Integration component within a broader AR educational platform. The purpose of this component is to support early literacy learning through a connected process that links student-generated sketches, English word formation, pronunciation, and AR-based

reward feedback. In the proposed workflow, the learner first draws a 2D sketch of a familiar object. The system then recognizes the sketch and uses the student's own drawing as the basis for creating a 3D representation, rather than simply retrieving a ready-made 3D object from a database. After that, the learner writes the English name of the object using air-writing and pronounces the same word aloud. When the spelling and pronunciation are correctly matched with the recognized sketch, the system generates a complete 3D model as an interactive reward using Unity and ARCore.

This approach is important because it transforms literacy learning from a passive and isolated activity into a multisensory and feedback-driven experience. Instead of learning drawing, spelling, and pronunciation separately, children engage with all three in a connected learning loop. By combining visual interaction, word formation, speech input, and AR feedback, the proposed system aims to improve student engagement, strengthen vocabulary association, and make literacy learning more enjoyable and effective.

This research also aligns with Sustainable Development Goal (SDG) 4: Quality Education, which promotes inclusive, equitable, and accessible learning opportunities for all. Since the proposed system is designed as a mobile-based and low-cost solution, it has the potential to be used in both classroom and home learning contexts. From a research perspective, the study belongs to the Mixed Reality (MR) research cluster and contributes to the areas of Augmented Reality, Computer Vision, Machine Learning, and Interactive Media Systems. By integrating these technologies into an educational context, the research aims to support the development of more intelligent and engaging literacy learning environments for primary school students.

1.2 Literature Review

Recent research in Augmented Reality (AR), computer vision, sketch understanding, hand gesture recognition, and speech recognition shows that interactive technologies can improve learner engagement and make educational experiences more meaningful for children. Studies on AR-based learning environments suggest that immersive and interactive digital systems can improve student motivation, attention, and conceptual understanding when compared with conventional teaching methods [1], [2]. However, the literature also shows that most existing approaches address only one part of the learning process at a time. Very few systems combine sketch-based interaction, gesture-

based word formation, pronunciation checking, and AR feedback within one unified educational workflow.

Augmented Reality in Education

AR has been widely studied as a teaching tool because it allows digital content to be placed within the learner's real environment. Radu's review of AR in education [1] highlights that AR can improve learner motivation, engagement, and conceptual understanding when compared with traditional media. Similarly, Billingham and Dünser [2] discuss how AR can create more interactive classroom experiences by linking digital content with physical learning contexts. These studies confirm that AR has strong educational value, particularly for young learners who benefit from visual and interactive learning experiences.

However, many AR systems mainly focus on visualizing content rather than analyzing learner-generated input. In many educational applications, AR is used to display pre-designed models, animations, or overlays, but not to validate what the learner has drawn, written, or spoken. Therefore, while AR is useful for engagement and visualization, its application to integrated literacy learning remains limited.

Sketch Recognition and Sketch-to-3D Approaches

Sketch-based interaction has been studied as a natural way for users to communicate with digital systems. Wang et al. introduced Sketch2Model [3], showing that freehand sketches can be transformed into 3D models. This is highly relevant to the proposed research because it demonstrates that human-drawn sketches can be used as a basis for digital 3D generation.

Other studies have investigated sketch-based shape retrieval and sketch recognition using machine learning methods [4]. These works show that 2D sketches can be interpreted computationally and linked with meaningful shape representations. In addition, advances in image recognition architectures and large-scale image datasets have improved the ability of systems to recognize shapes and visual patterns [10], [11], [12], [15].

Although these studies provide a valuable technical foundation, most of them focus on shape retrieval, general recognition, or technical modeling tasks rather than literacy learning. In many cases, the sketch is used only to retrieve a pre-existing 3D shape or

classify an object category, rather than being used in an educational workflow that connects the drawing with spelling, pronunciation, and learner feedback. Therefore, the literature supports sketch understanding as a technical possibility, but does not fully address the educational use case targeted in this study.

Air-Writing and Online Handwriting Recognition

Another important research area is air-writing and online handwriting recognition. Amma et al. [5] introduced an air-writing system for recognizing 3D-space handwriting, showing that written characters can be detected from motion trajectories rather than only from final visual shapes. Similarly, Chen et al. [6] demonstrated that sequential models can recognize online handwritten text effectively by using temporal writing information.

These studies are important because they show that writing recognition should not always rely only on the final written result. Instead, motion-based input can provide useful information for identifying letters and words. This is relevant to the proposed research because the learner is expected to form the English name of the recognized object through air-writing interaction.

However, many air-writing and handwriting recognition systems were designed for text entry, gesture input, or general human-computer interaction, rather than for literacy support for children. They often focus on recognition performance rather than educational guidance. As a result, although these systems demonstrate that air-writing is technically feasible, they do not by themselves provide a complete child-friendly learning solution.

Hand Tracking and Gesture Recognition

Research in hand tracking and gesture recognition has significantly advanced with the use of computer vision and deep learning. Molchanov et al. [7] showed that gesture recognition can be effectively performed using spatio-temporal learning methods. Similarly, Cao et al. [8] demonstrated real-time pose estimation, while Simon et al. [13] showed how hand keypoints can be detected from images with high accuracy. These contributions provide a strong foundation for detecting hand movement in real time using ordinary cameras.

For the proposed system, this research is useful because air-writing depends on reliable hand and finger tracking. If hand movement can be captured accurately, it becomes possible to interpret the learner's writing-related gestures and connect them with the expected word. However, most gesture recognition studies still focus on predefined gestures such as control commands or symbolic hand signs, rather than educational word formation tasks. Therefore, these techniques must be adapted carefully to suit phonics-related interaction.

Speech Recognition and Pronunciation Validation

Speech recognition is another key area relevant to the proposed research. Graves et al. [9] demonstrated that deep recurrent neural networks can significantly improve speech recognition performance. This is important for pronunciation validation because it shows that machine learning models can process spoken input accurately enough to support educational use.

For the proposed system, speech recognition is necessary because the learner is expected to pronounce the English word associated with the recognized sketch. Therefore, speech processing helps connect visual recognition, written word formation, and spoken pronunciation into one multisensory loop. However, many existing speech recognition systems are designed for general speech-to-text tasks or language technology applications rather than for integrated phonics learning. In many systems, pronunciation is treated as an independent task and is not directly linked with drawing, gesture-based writing, or AR-based rewards.

Summary of Existing Research and Limitations

The literature reveals several important insights. First, AR research strongly supports the value of interactive and immersive learning environments [1], [2], but most existing systems focus mainly on visualization. Second, research in sketch recognition and sketch-to-3D modeling shows that human drawings can be interpreted and transformed into digital forms [3], [4], but these approaches are rarely connected with literacy learning. Third, studies in air-writing, hand tracking, and gesture recognition confirm that motion-based writing input is technically possible [5], [6], [7], [8], [13], yet these systems are usually developed for general interaction tasks rather than child-centered educational applications. Finally, speech recognition research demonstrates that spoken

words can be accurately processed in real time [9], but it is often applied as a separate module rather than as part of a unified learning workflow.

Therefore, the literature suggests that the required technologies already exist in separate forms, but they are rarely combined into a single educational framework. There is still limited research on systems that connect student-generated sketch input, sketch-based 3D transformation, word formation through air-writing, pronunciation validation, and AR-based reward generation within one mobile learning environment. This limitation provides the foundation for the proposed Multisensory Phonics Integration component, which aims to bridge these isolated technologies and create a more engaging, connected, and feedback-driven literacy learning experience for primary school learners.

1.3 Research Gap

Although significant progress has been made in augmented reality learning systems, sketch understanding, air-writing recognition, hand tracking, and speech recognition, existing research and commercial applications still do not provide a complete solution for early literacy learning. The literature reviewed in Section 1.2 shows that most current systems address only one part of the learning process at a time, rather than integrating multiple learning modalities into a single educational workflow [1], [2].

One major limitation in existing literacy applications is that they usually separate drawing, writing, and pronunciation into different activities. Some applications focus only on tracing letters, while others emphasize pronunciation practice or AR-based visualization. As a result, young learners often do not experience a connected learning process in which visual recognition, written word formation, and spoken pronunciation reinforce each other. This separation reduces the effectiveness of multisensory learning, even though research suggests that learners benefit more when visual, motor, and auditory interactions are combined [1], [2].

Another important gap exists in the use of student-generated sketches within educational systems. Research in sketch-based modeling and sketch-to-3D transformation has shown that hand-drawn sketches can be interpreted computationally and used to generate digital 3D outputs [3], [4]. However, these studies are mainly designed for general modeling, shape retrieval, or technical interaction tasks, not for

early literacy learning. In many systems, sketches are used only to recognize an object category or retrieve a pre-existing 3D model. They rarely use the learner’s own drawing as the basis for a progressive educational workflow that connects sketching, spelling, pronunciation, and AR-based rewards.

Similarly, research in air-writing and online handwriting recognition confirms that motion-based writing input can be recognized effectively [5], [6]. Hand tracking and gesture recognition studies also show that finger and hand movements can be captured in real time using computer vision techniques [7], [8], [13]. However, these approaches are commonly developed for text entry, gesture control, or general human-computer interaction, rather than for literacy learning in primary education. They are rarely integrated with sketch-based object learning or speech-based vocabulary validation in a child-centered educational environment.

Speech recognition research has also advanced significantly, and spoken input can now be processed accurately enough to support educational applications [9]. Yet in most existing systems, pronunciation checking is implemented as a separate feature. It is not usually linked with learner-generated sketches, air-written word formation, or AR-based object feedback. Therefore, current systems do not provide a unified learning loop that connects what the learner draws, writes, and says.

These limitations indicate a clear research gap: there is currently no widely adopted mobile-based educational system that integrates student-drawn sketch understanding, sketch-based 3D generation, air-written English word matching, pronunciation verification, and AR-based reward feedback within one unified learning environment for primary school learners.

To address this gap, the proposed research introduces a Multisensory Phonics Integration framework within the broader AR educational system. In this workflow, the learner first draws a 2D sketch of a familiar object. The system recognizes the sketch and uses the learner’s own drawing as the basis for creating a 3D representation, rather than simply retrieving a ready-made 3D model. The learner then writes the English name of the recognized object using air-writing and pronounces the same word aloud. When the spelling and pronunciation are correctly matched with the sketch, the system generates a completed 3D model as an interactive AR reward using Unity and ARCore.

The novelty of this component lies in integrating student-generated sketch recognition, sketch-based 3D transformation, air-written word matching, pronunciation validation, and AR reward generation within a single mobile learning workflow. Unlike existing systems that handle these processes separately, the proposed approach creates a connected multisensory learning loop in which drawing, word formation, and spoken language are validated together. This makes the learning process more engaging, more interactive, and more suitable for early literacy development.

By combining augmented reality, computer vision, gesture-based interaction, and speech processing, the proposed research aims to contribute to the development of intelligent educational systems that improve literacy learning outcomes among primary school students.

Research Problem (Main)

How can a mobile-based Artificial Intelligence (AI) and Augmented Reality (AR) system be designed to support both geometry learning and English literacy learning in a single low-cost, real-time interactive educational environment for primary school learners?

Research Problem (Sub Component)

While the overall system addresses multiple educational functions, this individual research component specifically focuses on the **Multisensory Phonics Integration** module. Therefore, the sub-component research problem is defined as follows:

How can a mobile-based AR learning component be designed to recognize a student-drawn 2D sketch, transform that sketch into a 3D representation, verify whether the air-written English word and spoken pronunciation correctly match the recognized object, and provide a completed 3D model as real-time reward feedback for primary school learners?

Addressing this sub-component problem requires the design of a workflow that can connect sketch input, 3D transformation, spelling verification, pronunciation checking, and AR feedback into one unified learning process. Solving this problem will help bridge the gap between isolated learning technologies and a more connected, multisensory literacy learning experience.

2. OBJECTIVE

2.1 Main Objective

The main objective of this research is to design, develop, and evaluate a mobile-based multisensory phonics learning component using Unity and ARCore that can recognize student-drawn 2D sketches, transform them into 3D representations, verify air-written English word matching and spoken pronunciation, and provide interactive AR-based reward feedback to support early literacy learning among primary school students aged 6–12.

The proposed component aims to transform traditional passive literacy learning into an interactive augmented reality learning experience by enabling learners to complete a connected sequence of activities. First, the student draws a familiar object as a 2D sketch. The system then recognizes the sketch and generates a corresponding 3D representation based on the student’s drawing. Next, the learner writes the English name of the object using air-writing and pronounces the same word aloud. These inputs are processed using computer vision, word-matching logic, and speech recognition techniques to determine whether the learning task has been completed correctly.

To ensure that this objective is measurable, the system will be evaluated using performance indicators such as sketch recognition accuracy, accuracy of matching the air-written word with the recognized object, pronunciation validation accuracy, system response time, and user learning improvement measured through pre-test and post-test evaluations. The proposed component aims to achieve reliable recognition and validation performance while maintaining real-time responsiveness on standard Android mobile devices. By achieving these outcomes, the research seeks to demonstrate the effectiveness of integrating sketch-based interaction, pronunciation verification, and AR-based 3D reward generation in supporting multisensory phonics learning.

2.2 Specific Objectives

To achieve the main objective of the proposed multisensory phonics learning component, the following specific objectives are defined. Each objective addresses a key function of the system and is aligned with the identified research problem.

Table 1: Specific Objectives of the Research

Objective No.	Specific Objective
Obj 1	Develop a sketch recognition module capable of identifying student-drawn 2D sketches of familiar objects used in phonics learning activities. This objective focuses on enabling the system to interpret child-generated drawings accurately and map them to the correct object category so that the sketch can be used as the starting point of the learning workflow.
Obj 2	Develop a 2D-to-3D transformation process that converts the student's recognized sketch into a corresponding 3D representation. This objective focuses on using the learner's own drawing as the basis for generating a digital 3D form, rather than simply retrieving a pre-existing model, thereby making the learning experience more personalized and interactive.
Obj 3	Develop a word and pronunciation validation module capable of checking whether the English word formed through air-writing and the spoken pronunciation correctly match the recognized object. This objective focuses on strengthening vocabulary learning by connecting visual recognition, word formation, and spoken language within the same educational activity.
Obj 4	Implement an AR-based interactive reward mechanism using Unity and ARCore that generates the completed 3D object and related feedback when the learner successfully completes the sketch, word matching, and pronunciation tasks. This objective focuses on reinforcing correct learning behaviour through real-time visual and audio feedback in an engaging augmented reality environment.

3. METHODOLOGY

3.1 Research Approach

This research adopts a Design Science Research (DSR) approach because the study is focused on designing, developing, and evaluating a technology-based educational solution for a real-world learning problem. The identified problem is the lack of an interactive and integrated literacy learning system that can connect student-generated sketching, English word formation, pronunciation validation, and AR-based reward feedback within a single mobile learning environment. Since the research aims not only to study the problem but also to create a practical solution, Design Science Research is the most suitable approach.

The main artifact developed in this study is the Multisensory Phonics Integration component, which functions as a mobile-based AR learning module for primary school students. This artifact is designed to recognize a student's 2D sketch, transform the sketch into a 3D representation, verify whether the air-written English word and spoken pronunciation match the recognized object, and finally provide an interactive AR-based 3D reward. Therefore, the research is solution-oriented and involves both technical development and evaluation of the proposed component.

The Design Science Research process in this study is carried out in several stages. First, the problem domain is examined through literature review and analysis of existing educational systems. Second, the objectives of the proposed solution are defined based on the identified research gap. Third, the system architecture, interaction workflow, and validation process are designed according to the requirements of the multisensory phonics learning task. Fourth, the component is implemented using technologies such as Unity, ARCore, computer vision methods, and speech recognition techniques. Finally, the developed component is evaluated using both technical performance measures and user-centered learning outcomes.

This approach is appropriate because it allows the research to generate both a functional software artifact and an evidence-based evaluation of its usefulness in an educational setting. In addition, the DSR approach supports iterative refinement, meaning the component can be improved based on testing results and user feedback. As a result, the research does not only propose a theoretical concept, but also demonstrates a practical

and testable solution for improving multisensory phonics learning among primary school students.

3.2 System Design and Technical Methods

The proposed component follows a multimodal learning workflow in which different learner inputs are captured, processed, and validated through connected modules. The main purpose of this component is to support multisensory phonics learning by combining student-drawn sketches, sketch-based 3D generation, English word matching through air-writing, pronunciation verification, and AR-based reward feedback. Unlike conventional literacy tools that treat these tasks separately, this component links them into a single interactive learning process.

The system consists of four major functional modules: Sketch Recognition Module, 2D-to-3D Transformation Module, Word and Pronunciation Validation Module, and AR Reward Generation Module.

3.2.1 Sketch Recognition Module

The sketch recognition module enables the system to identify the object drawn by the learner. This module is important because the student's sketch acts as the starting point of the learning process. The learner is expected to draw a familiar object, such as an animal, fruit, or classroom-related item, and the system must recognize the drawing before moving to the next stage.

The sketch recognition process begins with image acquisition, where the mobile device camera captures the student's drawing input. The captured image is then passed through an image preprocessing stage, where operations such as grayscale conversion, noise reduction, background cleaning, and normalization are applied to improve the quality of the sketch. After preprocessing, the system extracts the main visual structure of the drawing, including lines, contours, and overall shape information. These features are then analyzed to determine the most likely object represented by the sketch.

Once the sketch is recognized, the system maps it to the corresponding English object name. This recognized label is then used in the next learning stages, including word matching, pronunciation verification, and 3D object generation.

3.2.2 2D-to-3D Transformation Module

After the object sketch is recognized, the next step is to transform the learner's 2D sketch into a 3D representation. This is one of the main distinguishing features of the proposed component. Instead of simply retrieving a ready-made 3D model from a database, the system attempts to use the learner's own drawing as the basis for generating a 3D form.

In this module, the recognized 2D sketch is analyzed to identify its outer structure and shape boundaries. The system then applies a transformation process to convert the 2D form into a digital 3D representation. This may include operations such as contour extraction, shape filling, extrusion, depth assignment, or geometric reconstruction depending on the type of object being handled. At the initial stage, the system may generate a partial or simplified 3D form, which will later be completed when the learner successfully finishes the remaining learning tasks.

This approach is more interactive and personalized because the resulting 3D output is connected to the learner's own sketch, rather than being a completely unrelated pre-designed object. Therefore, the module supports both creativity and learning engagement.

3.2.3 Word Validation Module

Once the sketch has been recognized, the learner is expected to write the English name of the object using air-writing interaction. In this component, the main focus is not on validating stroke order or detailed letter trajectory logic, but rather on checking whether the formed word correctly matches the recognized object.

The system receives the air-written letter sequence from the related writing component and compares the resulting word with the expected English label generated from the sketch recognition stage. In this way, the component performs word-level validation to confirm whether the learner has written the correct spelling of the object name.

This module is important because it connects object recognition with vocabulary learning. For example, if the learner sketches a cat, the system expects the air-written word to correspond to "cat." Therefore, the purpose of this stage is to ensure that the learner can correctly associate the drawn object with its English word form.

3.2.4 Pronunciation Validation Module

The pronunciation validation module evaluates whether the learner pronounces the English word correctly after completing the sketch and word formation stages. This module helps reinforce the connection between visual input, written language, and spoken language, which is central to phonics learning.

The process begins with audio capture, where the learner's voice is recorded through the mobile device microphone. The audio input is then preprocessed to reduce background noise and improve sound clarity. After preprocessing, the speech signal is analyzed using speech recognition techniques to identify the spoken word. The recognized word is then compared with the expected object name obtained from the sketch recognition stage.

If the spoken pronunciation matches the expected word with acceptable accuracy, the module confirms that the learner has completed the pronunciation task successfully. This stage is essential because it strengthens the learner's vocabulary recognition and spoken language skills within the same learning activity.

3.2.5 AR Reward Generation Module

When the learner successfully completes the sketch recognition, word matching, and pronunciation validation stages, the system generates an interactive AR-based reward. This module is implemented using Unity and ARCore and is responsible for presenting the completed 3D output as feedback to the learner.

At this stage, the earlier partial 3D representation is transformed into a more complete and visually meaningful 3D object. The generated object may be displayed with simple animation, motion, scaling, or sound-based reinforcement to create a gamified learning experience. This reward mechanism is designed to increase learner motivation and make the educational interaction more engaging.

Therefore, the AR reward generation module acts not only as a feedback mechanism, but also as a motivational learning tool that encourages children to complete the literacy task successfully.

3.3 Tools and Platforms

The proposed component will be developed using a set of software frameworks and tools selected based on their suitability for mobile-based AR application development, sketch processing, speech handling, and 3D content generation.

Unity 3D

Unity is used as the main development platform for implementing the user interface, interaction logic, and AR-based visualization of the system. It provides a flexible environment for integrating recognition modules with interactive 3D content and supports mobile deployment efficiently.

ARCore

ARCore is used to provide markerless augmented reality functionality on Android devices. It supports environmental understanding, surface detection, and real-time AR object placement, allowing the generated 3D object to be displayed within the learner's physical environment.

MediaPipe

MediaPipe is used as a supporting framework for real-time hand and finger tracking where required for the air-writing interaction process. It provides efficient landmark detection using standard camera input and supports mobile-friendly execution.

TensorFlow / TensorFlow Lite

TensorFlow is used for model development and training in recognition-related tasks such as sketch interpretation or supporting classification processes. TensorFlow Lite enables lightweight inference on mobile devices, making the system more practical for real-time use.

Blender

Blender is used to prepare, edit, or refine 3D assets and object representations used during AR visualization and reward presentation. It also supports prototyping of object forms when required for testing and comparison.

The combination of these tools allows the component to operate on standard Android devices while supporting computer vision processing, interactive feedback, and real-time AR visualization.

3.4 Data Collection and Ethical Considerations

The development and evaluation of the proposed component require data related to student sketches, word-level writing outcomes, and pronunciation recordings.

Data Sources

The data used in this study may include:

- hand-drawn sketches of familiar objects related to phonics learning
- air-written English word outputs corresponding to recognized objects
- voice recordings of students pronouncing those words
- supportive sample datasets from publicly available sources, where appropriate

Sketch samples may be collected from volunteer participants in order to understand how children naturally draw familiar objects. Word outputs and speech recordings may also be collected during controlled testing sessions to evaluate the accuracy of matching and pronunciation validation.

Ethical Considerations

Since the proposed research involves educational use and may include primary school learners during evaluation, ethical considerations are highly important.

The following ethical principles will be followed:

- participation in testing will be voluntary
- informed consent will be obtained from participants and, where necessary, from parents or guardians
- no unnecessary personal information will be collected
- collected sketches, writing outputs, and voice samples will be anonymized
- research data will be used only for academic and evaluation purposes
- the system will be designed to ensure safe and age-appropriate educational interaction

Approval will be obtained from the relevant institutional ethics review process before conducting user-based evaluation activities involving children.

3.5 Validation Strategy and Performance Metrics

The proposed component will be evaluated using both technical validation and user-centered educational evaluation.

Technical Performance Evaluation

The technical performance of the component will be measured using the following criteria:

Sketch Recognition Accuracy

This measures how accurately the system identifies the object drawn by the learner.

2D-to-3D Transformation Quality

This evaluates whether the generated 3D representation is meaningfully related to the learner's sketch and visually suitable for use as AR feedback.

Word Matching Accuracy

This measures whether the system can correctly validate the air-written English word against the recognized sketch label.

Pronunciation Validation Accuracy

This measures how accurately the system recognizes and validates the spoken pronunciation of the expected word.

System Response Time

This measures the time taken by the system to process inputs and produce AR-based feedback in real time.

User Experience and Learning Evaluation

To evaluate the educational usefulness of the component, user testing will be conducted with the target learner group.

Pre-test and Post-test Learning Assessment

Students will complete simple literacy-related tasks before and after interacting with the system, allowing measurement of improvement in word recognition, spelling association, and pronunciation understanding.

Usability Surveys

Students and teachers will provide feedback on system usability, clarity, engagement, and ease of interaction.

Observation-Based Evaluation

The researcher will observe how students interact with the system, including their level of interest, confidence, and response to AR-based rewards.

These evaluation methods will help determine whether the proposed component improves engagement, vocabulary association, pronunciation practice, and overall multisensory phonics learning effectiveness.

4. HIGH-LEVEL SYSTEM ARCHITECTURE

High-Level System Architecture of the Mobile-Based AI and AR Educational System for Primary Students

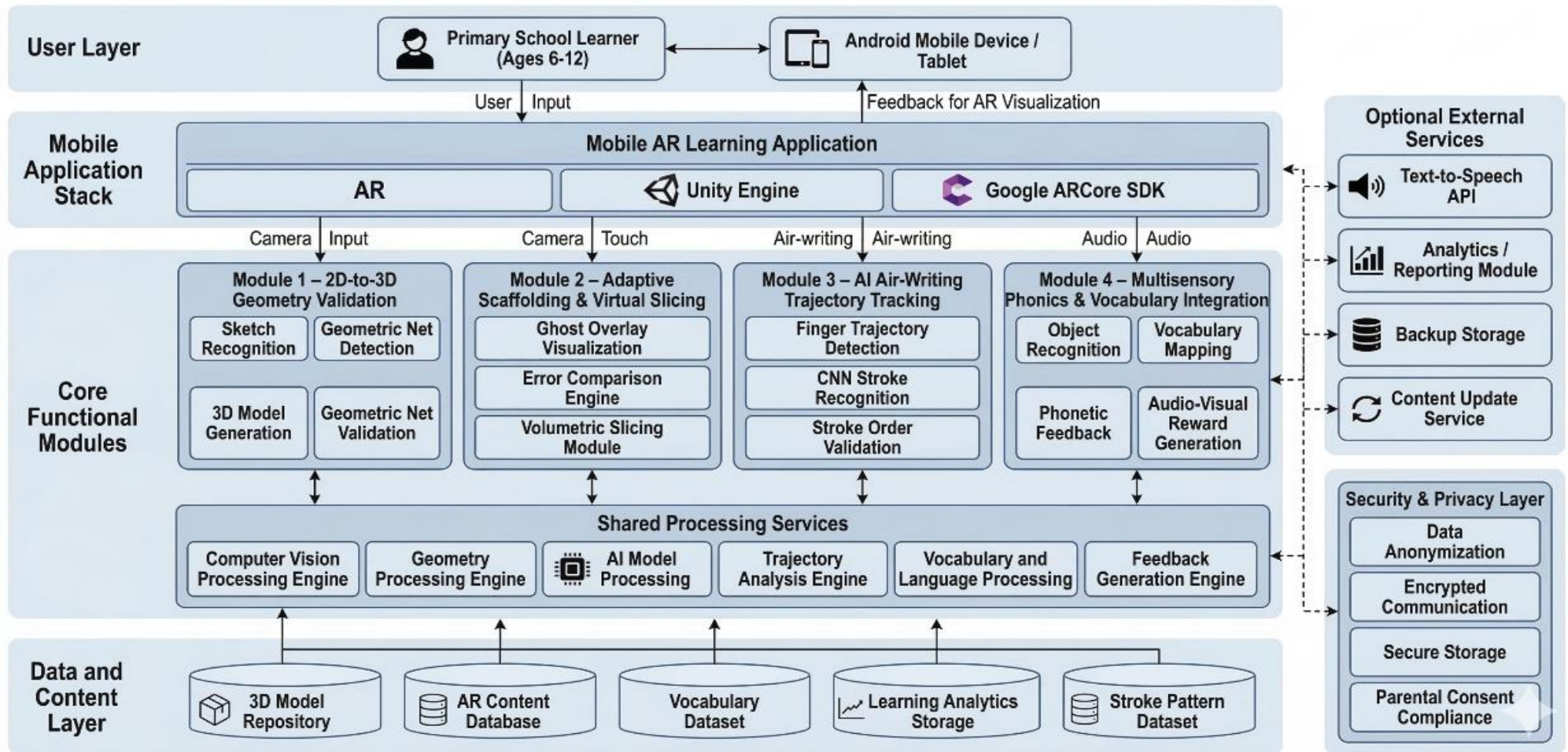


Figure 1: HIGH-LEVEL SYSTEM ARCHITECTURE

4.1 Overview of the Proposed Solution and Its Components

The proposed solution is a mobile-based Artificial Intelligence (AI) and Augmented Reality (AR) educational system designed to support both geometry learning and early literacy learning among primary school students. The overall project combines computer vision, gesture-based interaction, speech processing, and augmented reality to create a more engaging and interactive learning environment. Instead of relying only on traditional teaching methods such as textbooks and worksheets, the system allows learners to interact with educational content through drawing, movement, and speech while receiving real-time digital feedback.

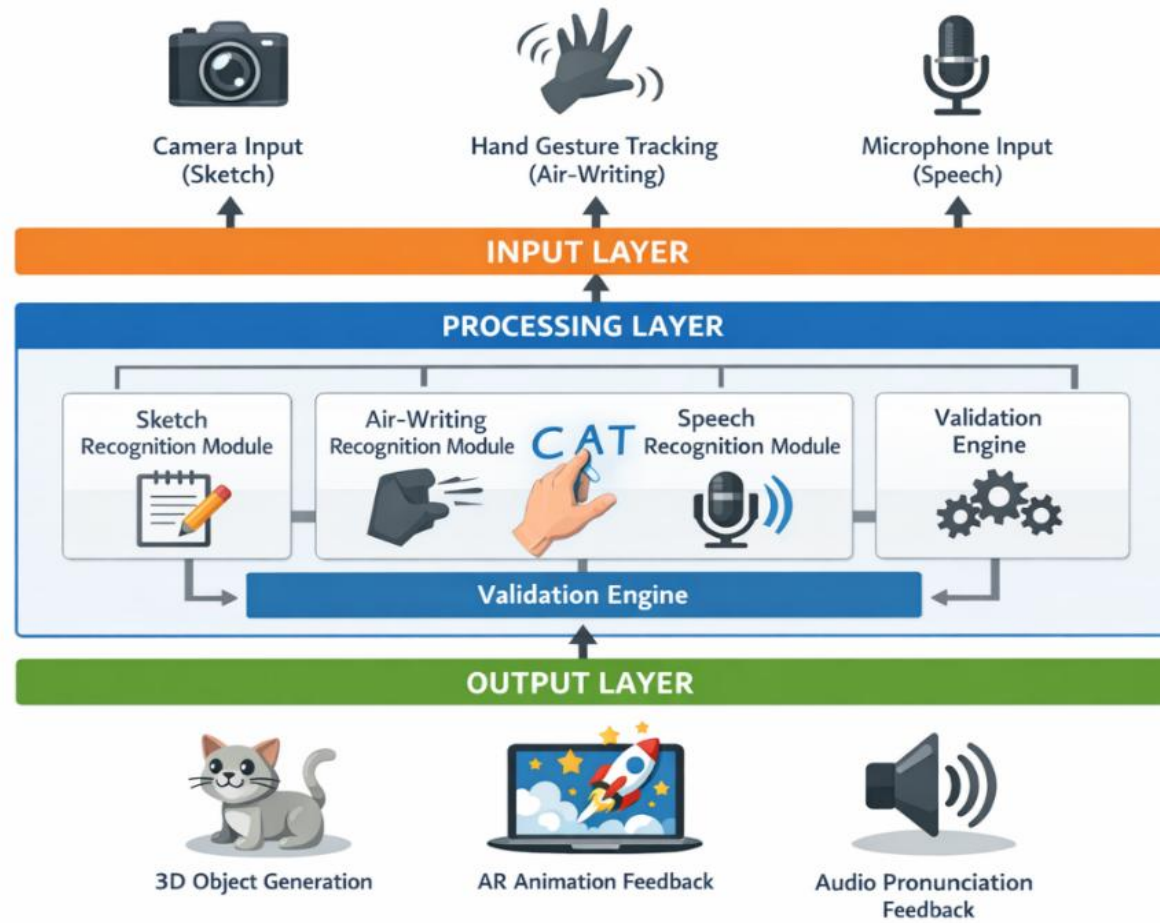
As a complete platform, the project includes several functional components that address different educational tasks within the broader AR Educational Suite. At the system level, these components work together to improve learning engagement, understanding, and participation. Some components focus on geometry-related tasks, while others focus on literacy-related tasks. Therefore, the overall project is not a single-purpose application, but a wider smart learning platform that uses multiple AI-driven modules to support primary education.

Within this broader project, the present individual research focuses specifically on the Multisensory Phonics Integration component. This component is designed to support literacy learning by connecting student-generated sketches, word formation, pronunciation, and AR-based reward feedback within one continuous learning workflow. In this process, the learner first draws a familiar object as a 2D sketch. The system then recognizes the sketch, transforms it into a 3D representation, and guides the learner to write and pronounce the related English word. When these tasks are completed successfully, the system provides interactive AR-based feedback as a reward.

Unlike conventional literacy applications that often treat drawing, writing, and speaking as separate activities, this component creates a multisensory learning loop where visual input, movement-based interaction, spoken language, and AR feedback are connected. This makes the learning experience more engaging and meaningful for children. In addition, the system is designed as a mobile-based and low-cost solution using Unity and ARCore, making it suitable for both classroom and home learning environments.

4.2 System Diagram of Components

Figure 2 : System



4.3 Individual Component and Its Parts

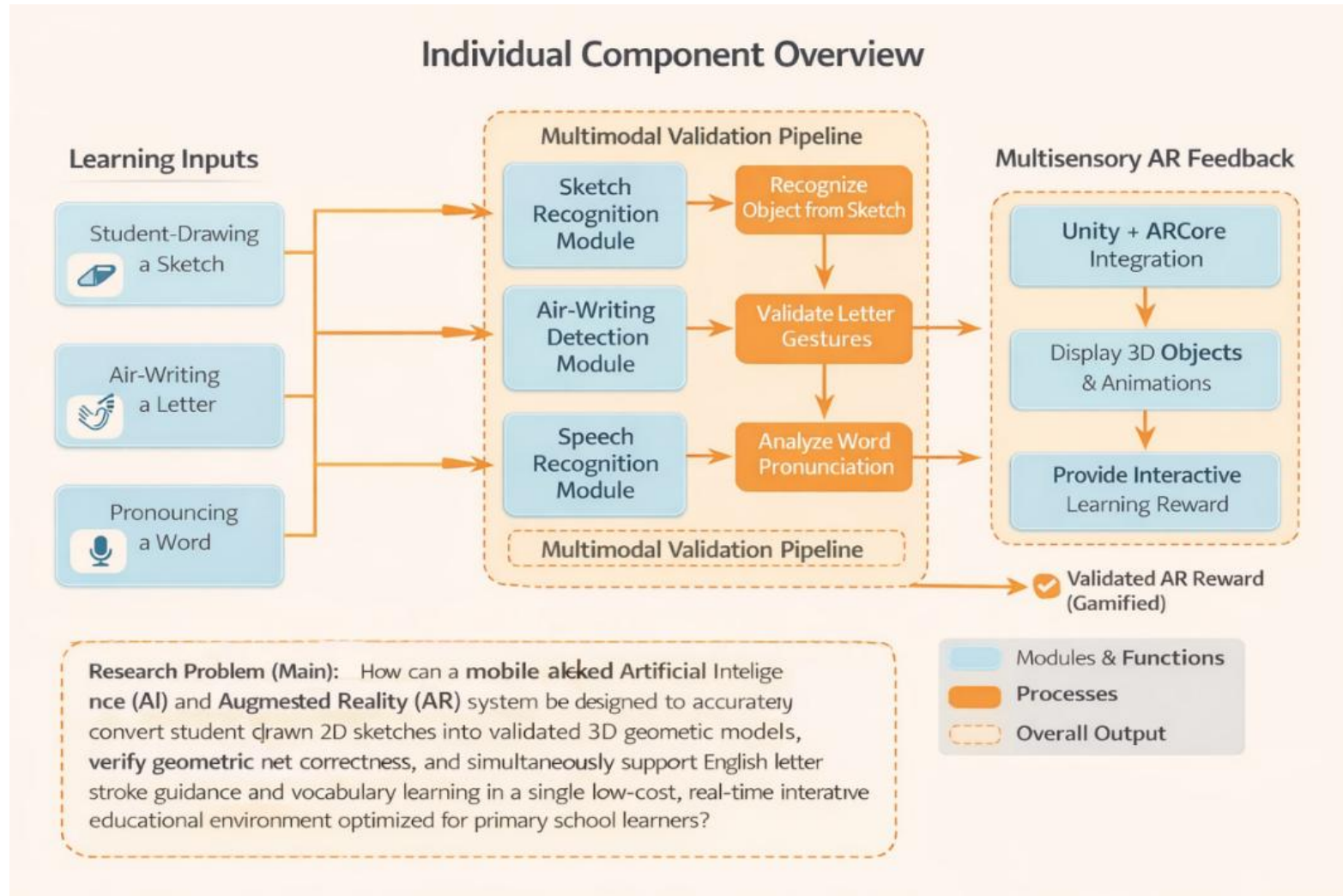


Figure 3: Individual Component and Its Parts

4.4 Processing Layer

The Processing Layer is the core of the proposed component. It consists of four main modules that work together to analyze learner input and validate the multisensory learning process.

4.4.1 Sketch Recognition Module

This module identifies the object drawn by the learner. The system captures the 2D sketch, preprocesses it, and analyzes its visual structure to determine the most likely object category. Once recognized, the object is mapped to its corresponding English word. This recognized label becomes the reference point for the next validation stages.

4.4.2 2D-to-3D Transformation Module

After the sketch is recognized, the system transforms the learner's own 2D sketch into a 3D representation. This is an important feature of the component because it does not simply retrieve a pre-made 3D model from a database. Instead, it uses the learner's drawing as the basis for generating a 3D form. At the initial stage, a partial or simplified 3D representation may be generated, which can later be completed after successful validation of the remaining tasks.

4.4.3 Word and Pronunciation Validation Module

This module checks whether the learner's air-written English word and spoken pronunciation correctly match the recognized object. At this stage, the focus of the component is on word-level matching and pronunciation validation, rather than detailed stroke-order analysis. The air-written result is compared with the expected English object name, and the spoken input is checked against the same word. This helps connect the learner's drawing with vocabulary formation and pronunciation practice.

4.4.4 Validation and Decision Module

This module acts as the central control point of the system. It receives outputs from the sketch recognition, 2D-to-3D transformation, word matching, and pronunciation validation modules. If the sketch, written word, and spoken pronunciation are all correctly aligned, the module confirms successful completion of the task and triggers the final AR-based reward stage.

4.5 Output Layer

The Output Layer provides feedback to the learner after the system processes and validates the input data.

The outputs include:

- **3D Object Visualization** – the learner’s sketch is shown as a completed 3D model
- **AR Reward Feedback** – the 3D object is displayed in the augmented reality environment using Unity and ARCore
- **Audio and Visual Reinforcement** – the system may provide supportive sound effects, spoken output, or animation as a reward mechanism

This layer is important because it transforms the system response into an engaging learning experience. Instead of only showing whether the learner is correct or incorrect, the system provides an interactive reward that motivates the learner to continue practicing.

4.6 Component Interaction and Data Flow

The data flow of the component begins when the learner draws a 2D sketch of a familiar object. The sketch is captured and analyzed by the Sketch Recognition Module, which identifies the object and maps it to its English label. The same sketch is then passed to the 2D-to-3D Transformation Module, where it is converted into an initial 3D representation.

Next, the learner writes the English name of the object using air-writing and pronounces the same word aloud. These inputs are processed by the Word and Pronunciation Validation Module, which checks whether both forms correctly match the recognized object. The results from all modules are then sent to the Validation and Decision Module. If all conditions are satisfied, the system sends a success signal to the Output Layer, where the completed 3D model is displayed as an AR-based reward.

This interaction creates a connected learning loop in which the learner’s drawing, word formation, pronunciation, and AR feedback are all linked together within a single workflow.

4.7 Scalability Considerations

The proposed architecture is designed to be scalable. Since each module is separated by function, the system can be extended in future work without redesigning the full architecture.

Possible future improvements include:

- adding more object categories and vocabulary sets
- supporting multiple difficulty levels
- expanding the system to other languages
- improving 3D generation quality
- integrating teacher dashboards or learner progress tracking

Because the system is implemented using Unity, ARCore, and mobile-compatible processing tools, it can be adapted for broader classroom and home-learning use.

4.8 Security and Privacy Considerations

Since the system may be tested with primary school learners, privacy and safety are important.

The architecture supports safe use through the following considerations:

- collecting only the minimum required sketch, word, and audio data
- anonymizing all evaluation data
- avoiding unnecessary storage of personal information
- restricting research data access to authorized users only
- ensuring age-appropriate educational interaction within the application

These measures help ensure that the component can be used responsibly in an educational setting.

4.9 Summary

In summary, the proposed high-level architecture is built around a three-layer structure consisting of Input, Processing, and Output. Within this structure, the Multisensory Phonics Integration component connects sketch recognition, sketch-based 3D transformation, word matching, pronunciation validation, and AR-based reward generation into one unified learning process. This architecture supports the main goal of creating an engaging, interactive, and feedback-driven literacy learning experience for primary school students.

5. USER REQUIREMENTS

5.1 Stakeholder Identification

Several stakeholders are expected to interact with, support, or benefit from the proposed Multisensory Phonics Integration component. Since the system is designed as a mobile-based AR learning solution for early literacy development, its impact extends beyond direct student use and includes educational and home-learning contexts.

Primary School Students

Primary school students aged 6 to 12 years are the main users of the proposed system. The component is designed to help them improve vocabulary learning, English word association, spelling, and pronunciation through interactive learning activities. Learners engage with the system by drawing familiar objects, producing the related English word, and pronouncing it aloud. Therefore, students benefit from a more engaging and multisensory learning experience than traditional literacy exercises.

Teachers

Teachers are important stakeholders because they can use the system as a supportive classroom learning tool. The component can assist teachers in introducing vocabulary and phonics activities in a more interactive way, while also helping them observe how students respond to sketch-based, word-based, and pronunciation-based learning tasks. As a result, the system can be used to strengthen literacy lessons and improve student participation during classroom activities.

Parents

Parents are also key stakeholders because the system can be used outside the classroom as a home-based learning support tool. Through mobile access, parents can encourage children to practice object recognition, English word formation, and pronunciation in a more enjoyable way. This provides additional learning opportunities at home and supports continued literacy development beyond school hours.

Educational Institutions

Schools, primary education centers, and other educational organizations are institutional stakeholders of the proposed system. These institutions may adopt the

solution as a digital educational tool to improve learning engagement and integrate interactive technologies into literacy teaching. The use of a low-cost mobile-based AR solution is especially beneficial for institutions seeking practical and innovative learning technologies for young students.

5.2 Functional Requirements

Functional requirements describe the specific operations and capabilities that the proposed Multisensory Phonics Integration component must perform.

FR1. Sketch Detection and Recognition

The system shall detect and recognize hand-drawn sketches created by students using the mobile device camera or drawing interface.

FR2. Object Identification

The system shall identify the detected sketch as a predefined familiar object related to literacy learning, such as an animal, fruit, or classroom object, and map it to the corresponding English word.

FR3. Sketch-Based 3D Generation

The system shall transform the recognized student-drawn 2D sketch into a corresponding 3D representation that can be used within the AR learning workflow.

FR4. Air-Written Word Input Reception

The system shall receive the English word produced by the learner through air-writing interaction for word-level validation.

FR5. Word Matching Validation

The system shall verify whether the air-written English word correctly matches the recognized object name obtained from the sketch recognition process.

FR6. Speech Input Capture

The system shall capture spoken word input through the device microphone during the pronunciation stage of the learning activity.

FR7. Pronunciation Validation

The system shall compare the spoken input with the expected English object name in order to evaluate pronunciation accuracy.

FR8. Multimodal Task Validation

The system shall combine the results of sketch recognition, word matching, and pronunciation validation to determine whether the learner has successfully completed the full learning task.

FR9. Augmented Reality Feedback

The system shall display the generated 3D object and related AR-based visual feedback when the learner successfully completes the required learning stages.

FR10. Gamified Learning Rewards

The system shall provide interactive reward feedback, such as 3D completion, animation, or audio reinforcement, to motivate learners and reinforce correct responses.

FR11. User Interaction Logging

The system shall record relevant interaction results, such as sketch recognition outcomes, word validation results, and pronunciation validation results, for testing and evaluation purposes.

5.3 Non-Functional Requirements

Non-functional requirements define the quality attributes, operational conditions, and system constraints of the proposed Multisensory Phonics Integration component.

NFR1. Performance Requirements

The system shall operate efficiently in order to provide a smooth and responsive learning experience for young users.

- The system shall process sketch, word, and speech inputs with minimal delay.
- The recognition and validation results shall be produced within a reasonable real-time response period after user interaction.

- The system shall maintain acceptable accuracy in sketch recognition, word matching, and pronunciation validation during normal usage.

NFR2. Usability Requirements

The system shall provide an interface that is suitable for primary school learners aged 6–12.

- The user interface shall be simple, child-friendly, visually engaging, and easy to navigate.
- Instructions and feedback shall be presented in a way that is easy for young learners to understand.
- The system shall guide the learner clearly when an input is incorrect or incomplete.

NFR3. Reliability Requirements

The system shall function consistently during repeated use under normal classroom or home-learning conditions.

- The application shall operate reliably during multiple learning sessions without frequent failure.
- The main modules shall provide stable outputs across different users and repeated attempts.
- The AR feedback mechanism shall function consistently once the learning task has been successfully validated.

NFR4. Security and Privacy Requirements

Because the system may involve child users, data privacy and safe usage are essential.

- The system shall avoid collecting or storing unnecessary personal information.
- Any sketches, word inputs, or voice samples collected for evaluation purposes shall be anonymized.
- Access to stored research data shall be limited to authorized researchers only.

NFR5. Compatibility Requirements

The system shall operate on commonly available mobile devices.

- The application shall run on Android devices that support ARCore.
- The system shall use standard device hardware, including the camera, microphone, and display, without requiring specialized external equipment.

NFR6. Scalability Requirements

The system shall support future improvement and extension without requiring complete redesign.

- Additional object categories, vocabulary sets, and phonics-related activities shall be able to be added in future versions.
- The recognition and validation modules shall be designed in a way that supports future refinement or replacement.
- The AR feedback module shall support expansion into richer 3D objects and more advanced reward interactions.

6. COMMERCIALIZATION PLAN

The proposed Multisensory Phonics Integration System has strong commercial potential as an educational technology product designed for early literacy development. By combining Augmented Reality (AR), computer vision, air-writing recognition, and speech recognition, the system provides an engaging and interactive learning experience for primary school students. This section describes the potential market, pricing strategy, cost estimation, revenue model, and intellectual property considerations.

6.1 Target Market and Customer Persona

Table 2: Target Market Segments

Target Market	Description	Potential Use of the System
Primary Schools	Primary schools in Sri Lanka represent the main institutional users of the system.	Schools can use the application as a digital teaching aid to support phonics learning, handwriting practice, and vocabulary development during classroom activities.
Educational Technology Companies	Local and international EdTech companies developing digital learning platforms.	These companies may integrate the system into online learning platforms, smart classrooms, or digital education tools .
Parents / Home Learning Users	Parents who want to support their children's literacy development outside school.	The application can be used as a home learning tool through smartphones or tablets to improve phonics, writing, and pronunciation skills.

Table 3: Customer Personas

Persona Type	Details
Teacher Persona	Name: Nadeesha Perera Age: 34 Occupation: Grade 2 Teacher Needs: Interactive tools that help students learn phonics and handwriting effectively Challenges: Limited time to provide individual feedback to each student Goal: Use digital tools that increase student engagement and improve literacy learning outcomes
Parent Persona	Name: Ravi Fernando Age: 40 Occupation: IT Professional Needs: Educational applications that help his child improve reading and writing skills Goal: Provide engaging and effective home-learning tools for his child

6.2 Value Proposition

The proposed system offers several unique benefits:

- **Multisensory Learning Experience** – Integrates visual, auditory, and motor learning through sketching, air-writing, and pronunciation.
- **Augmented Reality Engagement** – Interactive 3D animations improve student motivation and learning retention.
- **Real-Time Feedback** – Immediate feedback helps students correct mistakes in writing or pronunciation.
- **Gamified Learning Environment** – Students receive rewards and interactive animations when learning tasks are completed correctly.
- **Affordable Mobile Solution** – Works on standard Android devices using ARCore, making it accessible for schools with limited technological resources.

6.3 Revenue Model

The system can generate revenue through multiple channels.

School Licensing Model

Schools can purchase annual licenses to use the application across classrooms.

Estimated pricing:

Small schools: Rs. 30,000 – Rs. 50,000 per year

Medium schools: Rs. 50,000 – Rs. 100,000 per year

Large schools: Rs. 100,000+ per year

Mobile App Subscription Model

Parents can subscribe to the application for home learning.

Monthly subscription: Rs. 600 – Rs. 900

Annual subscription: Rs. 5,000 – Rs. 8,000

Educational Content Packages

Additional learning modules such as advanced phonics lessons and vocabulary activities can be sold separately.

Premium learning packs: Rs. 1,500 – Rs. 3,000

Institutional Partnerships

Partnerships with educational publishers, private tutoring centers, and EdTech companies can support large-scale distribution of the platform.

6.4 Cost Estimation

Table 4: Cost Category

Cost Category	Estimated Cost (LKR)
AR-compatible mobile testing device	Rs. 60,000
Cloud computing for machine learning training	Rs. 25,000
Software tools and plugins	Rs. 25,000
Testing and deployment	Rs. 30,000
Miscellaneous research expenses	Rs. 15,000

6.5 Pricing Strategy

The pricing strategy focuses on affordability and scalability.

Freemium Strategy

A basic version of the application will be offered for free with limited learning activities.

Premium Upgrade

Advanced AR lessons and additional content will require paid subscriptions.

Institutional Discounts

Schools purchasing licenses for multiple classrooms may receive discounted pricing packages.

This approach encourages user adoption while maintaining sustainable revenue generation.

6.6 Competitive Advantage

The proposed system provides several advantages compared with existing literacy learning applications:

- Integrated Multisensory Learning – Combines sketch recognition, air-writing detection, and speech recognition in one platform.
- Real-Time Learning Feedback – Immediate feedback improves learning accuracy and engagement.
- Augmented Reality Interaction – AR visualization enhances student motivation and participation.
- Gamified Learning – Interactive rewards encourage repeated practice.
- Low Hardware Requirements – Runs on standard Android smartphones without specialized hardware.

6.7 Intellectual Property Considerations

Software Copyright

The application's source code, user interface, and system architecture will be protected through software copyright.

Algorithm Protection

Machine learning models used for sketch recognition and air-writing detection may be treated as proprietary algorithms.

Trademark

The application name and branding can be registered as trademarks to establish a unique product identity.

Licensing Agreements

Educational institutions and partners will use the system through licensing agreements that define distribution and usage rights.

7. BUDGET AND JUSTIFICATION

The following table presents the estimated budget required for the development and testing of the proposed Multisensory Phonics Integration System. The costs are estimated in Sri Lankan Rupees (LKR) and reflect typical expenses required for mobile application development, machine learning experimentation, and AR content creation.

Table 5: Estimated Budget for the Research Project

Budget Category	Description	Estimated Cost (LKR)
Mobile Test Device	Android smartphone compatible with ARCore for testing the AR application	Rs. 60,000
Cloud Computing / Model Training	Cloud GPU resources for training machine learning models (sketch recognition and gesture recognition)	Rs. 25,000
Software Tools & Development Resources	Development tools, plugins, and possible API services	Rs. 25,000
Testing and Deployment Costs	Application testing, debugging, and prototype deployment	Rs. 30,000
Miscellaneous Expenses	Data collection, documentation, and research-related costs	Rs. 15,000
Total Cost		Rs. 155,000

Table 6: Budget Justification for the Research Project

Budget Category	Description / Justification
Mobile Test Device	A compatible Android device supporting ARCore is required to test the augmented reality features of the application. The device will be used to evaluate real-time gesture tracking, camera input processing, and AR rendering performance.
Cloud Computing Resources	Machine learning models used for sketch recognition and gesture classification may require GPU-based training environments. Cloud platforms provide scalable computational resources for training neural networks efficiently.
Software Development Tools	Although many development tools such as Unity, ARCore SDK, MediaPipe, and TensorFlow are available as open-source frameworks, additional libraries or plugins may be required during development.
3D Asset Creation	The system uses animated 3D objects as gamified rewards. These assets may be created using Blender or obtained from 3D asset marketplaces, which may involve licensing costs.
Testing and Deployment	Costs associated with application testing include debugging tools, usability testing sessions, and prototype deployment to mobile devices.
Miscellaneous Expenses	Additional expenses may arise during the research process, including data collection materials, printing, documentation preparation, and system testing activities.

8. WORK BREAKDOWN STRUCTURE (WBS)

Table 7: Work Breakdown Structure for Multisensory Phonics Integration Component

Task ID	Task Description	Timeline	Expected Output
1	Conduct literature review on sketch recognition, air-writing recognition, and speech recognition technologies for phonics learning	Month 1	Literature review document and research references
2	Collect and prepare datasets including hand-drawn sketches, air-writing gesture trajectories, and pronunciation recordings	Month 1 – Month 2	Prepared dataset for model training and testing
3	Design and develop the Sketch Recognition Module using computer vision and CNN models	Month 2 – Month 3	Trained sketch recognition model capable of identifying common objects
4	Implement the Air-Writing Detection Module using hand landmark tracking and gesture trajectory analysis	Month 3 – Month 4	Functional air-writing recognition system for detecting letters
5	Develop the Speech Recognition Module for phonics pronunciation validation	Month 3 – Month 4	Speech recognition module capable of validating word pronunciation
6	Integrate sketch recognition, gesture recognition, and speech recognition modules into a multimodal validation pipeline	Month 4 – Month 5	Integrated multisensory recognition system
7	Implement Augmented Reality feedback using Unity and ARCore to display 3D objects and animations	Month 4 – Month 5	Interactive AR learning interface
8	Perform system testing to evaluate recognition accuracy, latency, and performance	Month 5	System evaluation results
9	Conduct usability testing with target users and analyze learning effectiveness	Month 5 – Month 6	User testing report
10	Prepare documentation and research reports for proposal, progress presentations, and final submission	Month 1 – Month 6	Completed research documentation

9. GANTT CHART

The following Gantt chart illustrates the timeline for developing the Multisensory Phonics Integration component. The tasks are organized across six months to ensure systematic development, integration, and evaluation of the system.

Table 8:9. GANTT CHART

Task / Activity	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Literature Review								
Dataset Collection								
Sketch Recognition Model Development								
Air-Writing Recognition Module								
Speech Recognition Integration								
AR System Development (Unity + ARCore)								
System Integration								
Testing and Evaluation								
Improvements after Progress Presentation 2								
Final Documentation & Report Preparation								

10. REFERENCES LIST

- [1] I. Radu, “Why Should My Students Use AR? A Comparative Review of the Educational Impacts of Augmented-Reality,” in *Proc. IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Atlanta, GA, USA, 2012, pp. 313–314.
- [2] M. Billinghurst and A. Dünser, “Augmented Reality in the Classroom,” in *Proc. IEEE Int. Symp. Mixed and Augmented Reality (ISMAR)*, Basel, Switzerland, 2015, pp. 56–63.
- [3] X. Wang, K. Pulli, and J. Popović, “Sketch2Model: From Freehand Sketches to 3D Models,” in *Proc. IEEE Virtual Reality (VR)*, Singapore, 2011, pp. 61–68.
- [4] P. Xu, Q. Huang, and H. Wang, “Sketch-Based 3D Shape Retrieval Using Convolutional Neural Networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, USA, 2015, pp. 187–194.
- [5] C. Amma, M. Georgi, and T. Schultz, “Air-Writing: Hands-Free Mobile Text Input by Spotting and Continuous Recognition of 3D-Space Handwriting with Inertial Sensors,” in *Proc. ACM Int. Conf. Multimodal Interaction (ICMI)*, Santa Monica, CA, USA, 2012, pp. 52–59.
- [6] Y. Chen, F. Yin, and C. Liu, “Online Handwritten Text Recognition Using Recurrent Neural Networks,” in *Proc. IEEE Int. Conf. Document Analysis and Recognition (ICDAR)*, Nancy, France, 2015, pp. 1012–1016.
- [7] S. Molchanov, S. Gupta, K. Kim, and K. Pulli, “Hand Gesture Recognition with 3D Convolutional Neural Networks,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Boston, MA, USA, 2015, pp. 1–7.
- [8] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, “Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 7291–7299.
- [9] A. Graves, A. Mohamed, and G. Hinton, “Speech Recognition with Deep Recurrent Neural Networks,” in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, Canada, 2013, pp. 6645–6649.
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.
- [11] C. Szegedy et al., “Going Deeper with Convolutions,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, 2015, pp. 1–9.
- [12] A. Krizhevsky, I. Sutskever, and G. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proc. Neural Information Processing Systems (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

- [13] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand Keypoint Detection in Single Images Using Multiview Bootstrapping,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1145–1153.
- [14] N. Neverova, C. Wolf, G. Taylor, and F. Nebout, “ModDrop: Adaptive Multi-Modal Gesture Recognition,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 1692–1700.
- [15] J. Deng et al., “ImageNet: A Large-Scale Hierarchical Image Database,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255.

11.APPENDICES

The appendices include supplementary materials that support the proposed research methodology and system design. These materials provide additional details related to data collection instruments, system design artifacts, interface concepts, and project risk considerations.

11.1Appendix A – Survey Instruments

A short questionnaire will be used to collect feedback from teachers and students during the system evaluation phase.

Table 9: Observation Sheet

No	Question	Response Type
1	How effective are current phonics teaching tools used in your classroom?	Likert Scale (1–5)
2	Do students struggle with handwriting stroke order?	Yes / No
3	Would interactive AR tools help improve student engagement in literacy learning?	Likert Scale
4	How useful would real-time feedback on pronunciation be for students?	Likert Scale
5	What challenges do you face when teaching phonics and handwriting?	Open-ended

Student Feedback Questions

Table 10: Student Feedback Questions

No	Question	Response Type
1	Was the AR learning activity fun to use?	Yes / No
2	Did the system help you learn new words?	Yes / No
3	Was it easy to draw and write letters using gestures?	Likert Scale
4	Did the animations help you understand the word better?	Likert Scale
5	Would you like to use this system again?	Yes / No

11.2 Appendix B – Detailed Diagrams

The proposed system architecture consists of three main layers:

Input Layer

- Camera input for sketch detection
- Hand tracking for air-writing gestures
- Microphone input for speech recognition

Processing Layer

- Sketch Recognition Module (CNN-based)
- Air-Writing Trajectory Detection (MediaPipe + Gesture Analysis)
- Speech Recognition Module
- Multimodal Validation Engine

Output Layer

- AR visualization using Unity and ARCore
- Gamified reward animation
- Pronunciation feedback

11.3 Appendix C – Preliminary UI Mockups

The proposed mobile application will include the following main screens:

Table 11: Preliminary UI Mockups

Screen	Description	Key Features
Home Screen	Main entry interface where users start the learning activities and access system features.	• Start Learning • Progress Dashboard • Settings
Sketch Input Screen	Interface where students draw objects that will be recognized by the sketch recognition module.	• Canvas for drawing objects • Camera-based sketch capture • Recognition result preview
Air-Writing Screen	Screen that tracks finger movement to detect letter writing gestures in the air.	• Hand tracking visualization • Real-time stroke path display • Correct stroke guidance
Pronunciation Screen	Allows students to pronounce words while the system evaluates pronunciation accuracy.	• Microphone input • Pronunciation feedback indicator • Word pronunciation playback
AR Reward Screen	Displays interactive AR content as a reward when the student completes learning tasks correctly.	• 3D object visualization • Animation reward • Audio reinforcement

11.4 Appendix D – Risk Analysis

Table 12: Risk Analysis

Risk	Description	Mitigation Strategy
Recognition Accuracy	Sketch or gesture recognition may produce incorrect results	Improve dataset size and model training
Hardware Limitations	Low-end devices may struggle with AR rendering	Optimize models for mobile devices
Environmental Conditions	Poor lighting may affect camera tracking	Implement preprocessing and adaptive detection
Speech Noise	Background noise may affect speech recognition	Apply noise filtering techniques
User Interaction Complexity	Young students may find gestures difficult	Design simple and intuitive UI

11.5 Appendix E – Ethical Documents

This research will follow ethical guidelines to ensure participant safety and data privacy.

Key Ethical Principles

Informed Consent

Participants (students and teachers) will be informed about the purpose of the research and their participation will be voluntary.

Data Privacy

All collected data including sketches, gestures, and voice samples will be anonymized and used only for research purposes.

Child Protection

Since the research involves primary school students, permission will be obtained from school authorities and parents before data collection.

Data Security

All collected data will be securely stored and access will be restricted to the research team.