

Adaptive Learning Platform for English Language Development in Autistic Children in Sri Lanka

Adaptive Dialogue Learning and Progression Tracker

R26-IT-006

Project Proposal Report

M. C. Gunawardena – IT22166838

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

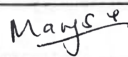
Department of Information Technology

Sri Lanka Institute of Information Technology Sri Lanka

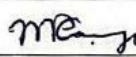
March 2026

DECLARATION

I declare that this is my own work, and this proposal does not knowingly incorporate any material previously submitted for a degree or a 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 with the text.

Name	Student ID	Signature
M. C. Gunawardena	IT22166838	

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

Name	Date	Signature
Ms. Manori Gamage	20/03/2026	

ABSTRACT

Autism Spectrum Disorder (ASD) affects approximately 1 in 127 children worldwide, with over 9,000 formally identified cases in Sri Lanka. Autistic children learning English as a second language face many barriers alongside communication deficits associated with ASD combined with the cognitive demands of acquiring a foreign language, making functional spoken dialogue one of the most difficult skills to develop. Existing technology-based interventions focus predominantly on isolated vocabulary recognition or assistive communication rather than structured dialogue learning. No reviewed system addresses pragmatic language understanding, generates dialogue constrained to a child's individual mastered vocabulary, or provides tolerant speech assessment suited to autistic non-native English speakers aligned to a national curriculum.

This research proposes the Adaptive Dialogue Learning and Progression Tracker, one component of AURIVA, a four-module adaptive English learning platform developed for Sri Lankan special education settings. The proposed module guides autistic ESL learners from vocabulary acquisition to functional conversation through three levels: dialogue word acquisition with pragmatic contextual understanding checks, personalized dialogue construction using a vocabulary-constrained Large Language Model, and sentence production with element-based partial credit assessment aligned to the Sri Lanka Grade 3 English curriculum. A contrastive instruction phase addresses the structural difference between Sinhala Subject-Object-Verb and English Subject-Verb-Object sentence order, with Sinhala instructional support used to reduce cognitive load.

The system is built using React Native, Node.js, Python Flask, spaCy, Google Cloud Speech-to-Text, the Google Gemini API, and fuzzball fuzzy string matching. A rule-based mastery detection algorithm governs progression, and a non-verbal fallback pathway ensures inclusive participation for minimally verbal learners. Validation targets include speech assessment classification accuracy of at least 80% and a session completion rate of at least 75% through a pilot study conducted at a partner special education school in Sri Lanka.

The proposed system directly contributes to SDG 4 Quality Education and SDG 10 Reduced Inequalities by providing a locally relevant, curriculum aligned, and accessible dialogue learning tool for a population that is significantly underserved by existing educational technology. Commercially, the platform targets Sri Lankan special education schools and ASD centres through a tiered institutional subscription model.

Keywords: Autism Spectrum Disorder, English as a Second Language, Adaptive Dialogue Learning, Speech Assessment, Large Language Model

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors for the continuous guidance and support provided throughout this research. I extend my appreciation to the staff and special educators at the partner special education school for their invaluable insights and cooperation during the requirement gathering process. I also thank my project group members for their collaboration throughout the proposal. Finally, I am grateful to my family for their encouragement and support throughout this journey.

TABLE OF CONTENT

Table of Contents

DECLARATION.....	2
ABSTRACT.....	3
ACKNOWLEDGEMENT.....	4
TABLE OF CONTENT	5
LIST OF TABLES.....	7
LIST OF FIGURES	8
LIST OF ABBREVIATIONS	9
LIST OF APPENDICES.....	10
1. INTRODUCTION.....	11
1.1 Background and Context.....	11
1.2 Literature Review	14
1.3 Research Gap.....	20
2. OBJECTIVE	26
2.1 Main Objective	26
2.2 Specific Objectives.....	26
3. METHODOLOGY	29
3.1 Research Approach.....	29
3.2 Technical Methods and Algorithms.....	30
3.3 Tools and Platforms with Justification.....	35
3.4 Data Collection and Ethical Considerations	37
3.5 Validation Strategy and Performance Metrics.....	38
4. HIGH-LEVEL SYSTEM ARCHITECTURE	42
4.1 Overview of proposed solution and its components.....	42
4.2 System Diagram of Components	43
4.3 Proposed Module – Component Interactions & Data Flow	44
4.4 Scalability and Security Concerns.....	45
5. USER REQUIREMENTS	47
5.1 Stakeholder identification	47
5.2 Functional Requirements	47

5.3 Non-Functional Requirements.....	48
6. COMMERCIALIZATION PLAN	50
6.1 Target market and customer persona	50
6.2 Value proposition	50
6.3 Revenue model	50
6.4 Cost estimation	51
6.5 Pricing strategy	51
6.6 Competitive advantage	52
6.7 Intellectual property considerations	52
7. BUDGET AND JUSTIFICATION	53
8. WORK BREAKDOWN STRUCTURE (WBS)	54
9. GANTT CHART	55
10. REFERENCES LIST.....	56
11. APPENDICES	58

LIST OF TABLES

Table 1 - Comparison of Existing Autism Communication Applications.....	19
Table 2 - Comparison of former research	21
Table 3 - The application of DSR in the research	29
Table 4 - Speech assessment pipeline breakdown	32
Table 5 - Tools and Platforms to develop the proposed module	35
Table 6 - Functional testing criteria for the proposed module	39
Table 7 - Speech assessment criteria for the proposed module.....	40
Table 8 - Educational Validity criteria for the proposed module	40
Table 9 - Proposed components of the system.....	42
Table 10 - Revenue model tiers	51
Table 11 - Budget and justifications.....	53

LIST OF FIGURES

Figure 1 - High level architecture overview of the system	43
Figure 2 - High Level architecture overview of the proposed component	44
Figure 3 - WBD structure.....	54
Figure 4 - Gantt Chart.....	55

LIST OF ABBREVIATIONS

Abbreviation	Description
AAC	Augmentative and Alternative Communication
AAC-BDAF	AAC integrated Big Data Analytics Framework
ABA	Applied Behaviour Analysis
ASR	Automatic Speech Recognition
ASD	Autism Spectrum Disorder
ASD-iLLM	Autism Spectrum Disorder Intervention Large Language Model
CNN	Convolutional Neural Network
ESL	English as a Second Language
GNN	Graph Neural Network
IEP	Individualized Education Plan
LLM	Large Language Model
MFCC	Mel-Frequency Cepstral Coefficients
NLP	Natural Language Processing
PECS	Picture Exchange Communication System
SGD	Speech Generating Device
STT	Speech-to-Text
SVO	Subject-Verb-Object
TTS	Text-to-Speech
UNICEF	United Nations Children's Fund
WHO	World Health Organization
UI	User Interface
UX	User Experience
JSON	JavaScript Object Notation
API	Application Programming Interface
AI	Artificial Intelligence
NGO	Non-Governmental Organization

LIST OF APPENDICES

Appendix 1- Level 1 Dialogue Word Acquisition Data Flow.....	58
Appendix 2 – Level 2 Dialogue Construction Data Flow.....	58

1. INTRODUCTION

1.1 Background and Context

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and interaction and restricted and repetitive patterns of behavior. Practical language competence, or the ability to use language functionally in various social contexts (e.g., to greet, request, or respond, and to maintain a conversation), is one of the most clinically-relevant areas of language development that is affected in children with ASD. For Sri Lankan autistic children learning English as a second language (ESL) these communication barriers clash with the mental strain of learning a new language, making spoken conversations one of the most difficult skills to develop in a conventional classroom setting.

The main groups affected by the lack of appropriate dialogue building language support are:

- **Autistic primary school children in Sri Lanka** who are expected to learn the national curriculum for English but struggle with basic functional communication. These children may be able to repeat sounds or recognize words but struggle to use the language appropriately in everyday interactions such as greeting, requesting, or responding to conversations.
- **Special educators** who are strained under the expectations to teach the given curriculum while accommodating diverse learning needs in children. Many rely on improvised methods or mainstream educational apps that are not designed for autistic children, which leads to inconsistent progress and high teacher workload.
- **Families and schools** face long-term impacts when children do not learn how to communicate effectively. The children may face social isolation, increase in behavioral challenges, limited academic achievement and lose the ability to gain further education or employment which lands more burden on the children's families.

As the English language is mainly used for education and employment in Sri Lanka, autistic children must be supported in acquiring basic dialogue skills as it directly contributes to education inclusion and equality.

According to the World Health Organization, (WHO) 1 in 127 children are affected by autism worldwide. According to a report by the Global Burden of Disease Study 2021 [14], the estimated number of people living with autism worldwide is around 61.8 million. In 2022, the United States Centers for Disease Control and Prevention (CDC) [15] reported that the autism prevalence among 8-year-old children was 1 in 31, up from 1 in 36 in 2020 and 1 in 44 in 2018. Autism is now included in the list of the top ten causes of non-fatal health burden in the under 20 years age group. A 2023 WHO-UNICEF report [16] estimated that 317 million children and young people globally live with health conditions contributing to developmental disabilities, many of whom face discrimination, educational exclusion, and

barriers to healthcare. In Sri Lanka, a 2009 local study [17] estimated an autism prevalence of 1.07% (approximately 1 in 93 children aged 18-24 months) and as of 2025, the Ministry of Women and Child Affairs said that over 9,000 children were formally diagnosed with autism, but more children were yet to be identified as most screenings have not taken place due to insufficient facilities [18]. There are no population-based screening programmes for autism in Sri Lanka and screening of Autism is not a part of primary health care. In addition, there is a lack of speech and language therapists, as nearly all graduates are employed in hospitals or the private sector, rather than the public school system.

Existing research has explored interventions using technology for autism communication support including tablet-based applications, Augmentative and Alternative Communication (AAC) systems, social robots, and large language model (LLM) based dialogue systems. Comprehensive reviews show that using technology for guided vocabulary learning can yield positive outcomes in receptive and expressive vocabulary for autistic children. However, the current work on speech therapy assistive systems all shares some weaknesses. Reliance on pre-built fixed wordlists instead of a learnt mastery profile for the children. The absence of verifying the children's understanding of pragmatics as no evaluation is conducted to assess in which situations the child learned how to use the vocabulary items, as learned by the model. Speech analysis technologies used as building blocks are rarely adapted to autistic children and non-natives, who frequently use non-standard pronunciations. Recent LLM based systems such as ASD-iLLM enables to generate relevant conversations from any word taken from the vocabulary of the intended language, whereas robot-assisted speech therapy applications currently have either fixed learning curricula or require the intervention of a human trainer or interpreter at every step of the activity, and AAC applications such as AAC-BDAF aim at the assistive aspect of choosing symbols that will represent the intent of the child in written language using text-to-speech techniques, while omitting the conversation or dialogue functionality, which again points towards the necessity of developing speech therapy assistive systems that include the following essential features for the benefit of children with autism, in particular functional dialogue teaching, pragmatic context-based evaluation, vocabulary- constrained personalization, speech evaluation with some level of tolerance towards pronunciations given by autistic and non-native children, as well as curriculum-tracking functionalities, all seamlessly integrated into one supportive learning system.

The proposed project aims to bridge this gap by designing and implementing an innovative learning system of 4 modules, to cater to the requirements of teaching English Language to children with autism in Sri Lanka up to Grade 3 of the National Curriculum. The proposed component is the dedicated component which deals with functional spoken dialogue. This has three levels: (1) **Dialogue Word Acquisition** with a three-stage learning cycle of Introducing words that are needed to be learnt and the greetings and phrases of politeness using a three-phase process of Familiarization, Production and Contextual Understanding. (2) **Personalized Dialogue Construction** where an LLM (Gemini API) generates unique conversation turns limited to the child's mastered vocabulary and validated against the learner profile before delivery to the child. (3) **Sentence Production with Partial-Credit Assessment** which is designed to break down the tasks of the Grade 3 syllabus (to introduce

oneself, to describe a friend, to describe a pet) into communication elements and scores the achievement of individual elements rather than a pass or fail system. In addition to this, the speech assessment function of the proposed project has a multi-layered approach using algorithms for the empty-check detection, keyword matching and fuzzy string matching to deal with issues of possible mispronunciation by the learner as well as speech to text errors. The non-verbal fallback flow of the assessment function that deals with non-speaking children provides the child with image-based and word matching activities.

This research aligns with **Sustainable Development Goals number four and five**. The SDG 4, which is **Quality Education** calls for inclusive and equal quality education and promote lifelong learning opportunities for all, eliminate disparities in and of quality education, ensure that all learning processes are inclusive for all, especially those with disabilities and children in vulnerable situations. By providing a curriculum aligned, dynamic, digital communication aid for autistic children, the proposed module directly contributes to inclusive primary education. The research will further contribute to achieving SDG 10, **Reduced Inequalities** within and among countries by reducing educational inequality experienced by autistic children in Sri Lanka due to the lack of specialist support services, trained teachers and suitable assistive technology. A functional communication skill-building, locally relevant and low resource friendly tool will contribute to reducing these inequalities and will improve the social opportunities of the autistic children in Sri Lanka thereby increasing their opportunities for schooling and future employment. This proposed initiative is developed under the research cluster of **Software Systems and Technologies (SST)**, targeting to cater to the needs of Education and Technology industry verticals. The work addresses fundamental SST concerns including system architecture design, real-time data processing, API integration, and intelligent content generation within a scalable software solution.

1.2 Literature Review

AURIVA is a four-module adaptive English learning platform developed for autistic children in Sri Lanka, aligned to the government primary English curriculum for Grades 1 to 3. The system is designed as an integrated mobile application where each module addresses a distinct dimension of English language learning and shares a common learner profile database. The **GNN-Enhanced Adaptive Concept Learning Module** teaches English vocabulary and attributes through a three-tier architecture of visual familiarization, word recognition, and attribute introduction, using a Graph Neural Network over a knowledge base to detect confusion patterns and personalize learning paths. The **Motor-Informed Adaptive Handwriting Learning Module** captures handwriting stroke data through a tablet interface; analyses trajectory and motor comfort scores using a rule-based evaluation model and adapts the letter learning sequence to each child's motor ability. The **Adaptive English Pronunciation Support Module** supports foundational English pronunciation through multimodal practice with images, audio and lip-sync guidance, using MFCC feature extraction and Dynamic Time Warping to score pronunciation similarity against a UK-accent reference audio library. The **Adaptive Dialogue Learning and Progression Tracker**, which is the focus of this research, teaches functional spoken dialogue through three levels that are greeting and magic word acquisition, personalized conversation construction using a vocabulary-constrained LLM, and sentence production with partial credit assessment aligned to the Grade 3 curriculum.

Literature Survey

The basis of using digital technologies for social communication in autism was explored by Jiang and Cohen [1] where a systematic review of 31 peer reviewed articles examined how digital technologies support social communication skills in autistic children and adolescents. The technologies reviewed were assigned into categories based on their functions which included virtual reality (VR) (n=13), augmented reality (AR) (n=5), touchable screen technology (e.g., iPads, tablets, and Surface devices) (n=3), computer-based programs (n=1), video modeling (VM) (n=4), a combination of computer-based programs and video modeling (n=2), AR combined with VM (n=1), interactive video games (n=1), and robots (n=1). The skills targeted with the use of the technologies included responding and maintaining conversations, initiation skills such as greetings, emotion recognition, and nonverbal communication. VR emerged as the technology with most promise offering controlled, predictable environments for repeated social practice. As a conclusion, the review showed that no single digital tool can address all social communication skill areas, indicating a clear need for tailored, multi-faceted interventions. A combination of technologies (e.g., AR with video modeling) yielded stronger outcomes in some studies. However, there were significant limitations to note; most studies relied on small, uniform samples; few tests were made on the

technology's applicability in real world settings and the longitudinal data for skill maintenance remains scarce. It is important to note that none of the 31 studies examined pragmatic language use; essentially the understanding of when and why a phrase is used in context which is the core novelty proposed in proposed module.

Farzana, Sarker, Chau, and Mamun [2] reviewed 60 research articles on the technological evolution of Augmentative and Alternative Communication (AAC) modalities for verbally challenged ASD children. The paradigm shifts from Picture Exchange Communication Systems (PECS) to Speech Generating Devices (SGDs), and later to mobile applications integrated with AI, ML, and AR were traced in this review. 53.33% of the articles reviewed were on mobile applications with the key features highlighted as customization, multimedia presentation, and progress report generation. AI/ML based applications such as LIVOX demonstrated how intelligent algorithms could reduce cognitive and motor demands on ASD children by predicting and adapting communication options. However many of the systems were designed for developed countries, which failed to address the cultural and linguistic needs of lower to middle income countries. The recommendations given in the review were; integrating intelligent analytics, personalization, and progress monitoring into a central proposed system. A critical gap to be noted is that while many applications focused on requesting skills or single word expression, none integrated multi-level dialogue building with adaptive progression tracking across interconnected learning modules.

Cole, van Vuuren, Pellom et al. [3] explored the use of animated agents and dialogue-based learning with the foundational research on perceptive animated interfaces and described experiments with an animated 3D talking head called "Baldi" that taught speech and language skills to profoundly deaf children and children with ASD. A four-stage sequence was followed by vocabulary tutors: pretest, introduction, practice, and post-test. The results revealed that children learned vocabulary quickly, retained over 50% of words months later, and perceived Baldi as a personal coach. For ASD children, six participants (ages 3–13) attending bi-weekly sessions for eight weeks demonstrated rapid concept acquisition, with teacher reports confirming applicability to classroom settings. Although foundational, this work established that even a primitive animated agent which is limited to mouse click interactions without speech recognition could engage autistic children and produce measurable learning gains. The absence of adaptive dialogue capabilities was a key limitation; the system could not dynamically adjust content based on the learner's progress. The proposed module addressed this gap through speech capture, fuzzy matching for assessment, and adaptive scaffolding driven by each child's performance profile.

Kumar, Balamurugan, Suriya, and Reshma [4] proposed an AAC integrated Big Data Analytics Framework (AAC-BDAF) for English language learning among children with neurological disorders. This framework adopted eye gaze tracking via Tobii X2-60 hardware to capture learner attention, with a CNN based prediction model achieving 98.63% accuracy in identifying selected options from gaze data. The predicted choices were converted to audible output with the use of a text to speech module. The framework demonstrated the effective integration of eye tracking with predictive analytics for passive learners focusing

solely on receptive identification through eye gaze rather than active speech production or conversational dialogue building. Additionally, resource limited school environments like those in Sri Lanka cannot rely on specialized Tobii hardware as it limits scalability, whereas the proposed module operates entirely on standard mobile devices.

The review on literature based on adaptive scaffolding and personalized learning is evaluated by Vogt et al. [5] where an adapt and explain strategy was evaluated based on a social robot which dynamically adjusted behavior and tasks based on the cognitive and affective of children aged 4-7 during second-language learning. This evaluation with 40 children highlighted that adaptive learning had a positive effect, with verbal explanations of the adaptations being particularly beneficial for slow learners. The robot successfully re-engaged children who had started to disengage in 79% of the cases, helping them achieve higher learning gains. This finding directly supports the adaptive guided teaching approach in the proposed module, where the system detects struggle patterns of a child and automatically revisits earlier phases with new activities and contextual scenarios. The proposed system delivers the equivalent adaptive scaffolding through a mobile app with an animated avatar without the limitation of the physical robot hardware which is significantly more accessible and scalable.

A systematic review by Urrea, Fernández-Torres, Rodríguez-Ortiz, and Saldaña assessed 13 studies published between 2006 and 2022 on technology-assisted vocabulary interventions for children with ASD. Five studies found positive outcomes, six found mixed outcomes, while one study found negative outcomes, and one study found no differences. The findings also highlight the importance of personalized assessment, knowledge activation, and contextualized learning as important factors in technology-mediated vocabulary instruction, which supports the customized learning in the proposed module by varying the vocabulary and dialogue level based on the concepts that each child has learned from the related modules. Thus, the activities are kept within the child's zone of proximal development.[6]

Aburukba, Aloul, Mahmoud, Kamili and Ajmal [7] created the AutiAid learning mobile application to assist autistic children, which includes a memory card game, a math and alphabet teaching game, scheduling and location monitoring. The application was tested at an Autism Centre in the UAE, which highlighted that memory and mathematics skills improved with repeated use of the application. The data-centered architecture eased data sharing between caregivers and autistic children. However, AutiAid did not address language learning or dialogue-building skills, and was designed around games with discrete learning objectives instead of multi-phase task learning objectives. This module builds on AutiAid's progress tracking concept while including a focus on conversational English learning with multi-dimensional mastery gating.

The use of LLM-based dialogue generation for autism was explored by Mishra, Welch, and Popa where they studied the use of LLMs for increasing robot autonomy in ASD interventions, using fine-tuned GPT-2 and BART models on the SOCIALIQA dataset to generate social scenarios and questions. The roles of initiator, prompter, and reinforcer were assumed by the robot with a human mediator validating appropriateness before delivery to

the child. Domain experts found that robot-assisted sessions did not significantly increase mental workload, and the robot was perceived as safe and likeable. This method of executing validation with a human mediator parallels the proposed module's approach, where a teacher provides child-specific information and the system's validation layer checks LLM-generated dialogue turns against the learner profile before presentation. The limitation is that this study used physical robots and was not tested with actual ASD children in naturalistic settings.[9]

Lai, Li et al. [8] proposed ASD-iLLM, a framework for training Large Language Models to conduct dialogue interventions for autistic children based on Applied Behavior Analysis (ABA) principles. The researchers built a multi-turn dialogue dataset (ASD-iLLM-8k) from clinical recordings and fine-tuned LLMs that outperformed existing models in both automatic and human evaluation, closely resembling clinical intervention dialogue styles. This is highly relevant to the proposed module's Level 2 (Dialogue Construction), which uses an LLM to dynamically generate personalized conversation turns constrained by each child's mastered vocabulary. However, ASD-iLLM targeted clinical therapeutic dialogue in Mandarin rather than structured English language teaching and lacked a validation layer to verify generated content against a learner's known vocabulary profile.

Speech recognition and assessment are covered in the review where Bosker [10] evaluated fuzzy string matching methods for automatic computation of speech intelligibility ratings. Results showed that Token Sort Ratio metric was highly correlated to human ratings ($r=0.940$). This analysis indicates that the fuzzy string-matching Fuzzball library to be used in the proposed module is appropriate for matching child speech transcriptions to expected child transcriptions with an 80% string-similarity threshold, accounting for speech mispronunciations that young, non-native English speaking children with ASD tend to make. Fuzzy matching tends to better match communicative intent and practice production and encourages a child to get credit for trying to produce the specified phrase rather than match pronunciation.

Tran, Son Le, Hung and Nguyen [11] tested speech recognition accuracy for non-native children, and Google Speech-to-Text made major mistakes due to non-native accents. For example, Vietnamese children aged 7-10 speaking English had "Do you have a smartphone?" transcribed as "Do you have a microphone?". Pham and Nguyen proposed an error correction and improvement method that is based on Bag-of-Words and deep neural network approaches. These methods are relevant to the proposed module that seeks to improve Sri Lankan second-language English speaker accuracy. A pipeline of empty check, keyword matching, and fuzzy matching is a reasonable practical error mitigation strategy, although the specific 80% threshold would need empirical calibration through testing with the relevant population of users.

The intersection of autism spectrum disorder and English as a second language learning presents a specific and underexplored challenge that directly shapes the design of the proposed module. Hashim, Yunus and Norman [12] investigated the obstacles faced by autistic learners in ESL settings through observations and interviews at two autism centers in Malaysia involving forty-five learners and fourteen teachers, finding that the world of ASD is

full of obstacles in ESL learning, yet not much spotlight or acknowledgment has been given to learners with autism in their journey of English language learning . The study identified communication deficits, sensory processing challenges, and the difficulty teachers face in developing appropriate educational materials as the primary barriers. Crucially, the findings confirmed that autistic ESL learners require highly structured, repetitive, and visually supported instruction which are characteristics that are built into every phase of the proposed module's design.

Mohamed and Shaaban [13] reviewed studies at the intersection of ASD and ESL, finding that teaching English as a second language to children with autism spectrum disorders using visual aids, modelling, and repetition has been shown to be an effective method to enhance language and social abilities, and that students with ASD can benefit significantly from ESL learning when educators recognize and address their unique needs. The review further noted that computer-based instruction, instructional software, and interactive multimedia can offer customized multimodal learning experiences that enhance language acquisition for this population. This directly supports the proposed module's use of avatar modelled audio, contextual scene images, and adaptive scaffolding as the primary instructional mechanisms. However, neither of these studies addressed sentence structure awareness or the structural interference caused by the learner's first language which is a gap that the proposed contrastive SVO (sentence, verb, object) instruction phase in Level 3 is specifically designed to address.

Existing Research Approaches

Four dominant approaches for technology assisted communication and language learning for autistic children are revealed through the literature. The majority is **single technology interventions**: Jiang and Cohen's [1] review revealed that VR was used in 13 of 31 studies, with the main target being conversation and initiation skills, but the lack of multi technology integration limits the effectiveness across the full spectrum of communication skills. **AAC-based communication aids** (PECS, SGDs, symbol-based mobile apps) do not support dialogue building; conversations with turn taking or contextual understanding of when to use specific phrases but succeeds at functional communication. **ABA-based structured learning** applications were successful at demonstrating their effectiveness for receptive language skills, and ASD-iLLM [8] extends this to LLM-driven dialogue interventions, but these methods do not combine with cross module vocabulary awareness or adaptive progress tracking. **Adaptive Guided Learning Systems** establish that real-time engagement tracking and dynamic content adjustments improve the learning outcomes and trajectory of struggling learners, although it has not been applied specifically to structured English language conversation building and learning for children with autism. [2][5][6]

Existing Products and Services

Table 1 - Comparison of Existing Autism Communication Applications

No table of figures entries found.	Type	Key Features	Limitations
Proloquo2Go	Symbol based AAC	Customizable vocabulary, natural voices, supports beginner to advanced users	Only available on iOS devices. Focuses on functional communication; no conversation building or adaptive progression tracking
Speech Blubs	Speech learning app	Child avatar teaches sounds, words, sentences and face filters for engagement	Speech imitation focus; no cross-module integration or vocabulary aware adaptive dialogue
Mita	AI powered language therapy	Dynamic content generated for the child's age, Difficulty of activities adapt with the child's learning trajectory	Targets receptive skills and conceptual understanding; no dialogue construction or contextual pragmatic use
Avaz	Spoken AI powered AAC	Predictive text, natural voices, learns from user patterns	Assistive communication tool; no educational guided learning or dialogue building
TappyTots	Button-based AAC	Child progress tracking, functional communication of various topics using buttons	AAC tool for basic needs; no conversational practice, speech capture, or adaptive learning

Although the existing products prove that the current market offers effective AAC and receptive language tools, there is no product that combines adaptive dialogue word learning, LLM constrained conversation building, speech-based production practice and contextual awareness for words and phrases within a single integrated system.

Limitations of Existing Systems

Existing systems do not integrate dialogue word learning with contextual awareness for pragmatic understanding. Current tools either teach vocabulary separately or focus on social communication without building structured conversation. The proposed three phase approach with word familiarization, production and contextual awareness check; bridges the

vocabulary learning with pragmatic language use in a unique manner, consistently testing as to whether a child understands when and why to use a word or phrase.

Integration of cross-module adaptive vocabulary is absent. The reviewed systems do not dynamically generate learning content based on the child's mastered knowledge from interconnected modules. The proposed Level 2 of the Adaptive Dialogue Learning and Progression Tracker component pulls live data from the system's other modules to build personalized conversations using only the base concepts and words the child has mastered, where the given learning content and activities scale automatically with the child's growing vocabulary.

Using an LLM to generate conversations constrained to the child's mastered vocabulary is unexplored. While ASD-iLLM demonstrated LLM-based dialogue interventions and other studies used LLMs for social cognition tasks, there is no review of using LLM's to dynamically generate an English language conversation by limiting the output to the child's learnt knowledge with a validation layer to ensure instruction design appropriateness.

Assessment of speech production in non-native autistic children lacks practical solutions. Rather than relying on exact matching in the speech produced by the child, the proposed three layer assessment pipeline offers a learning middle ground that focuses on learning intent rather than pronunciation perfection.

Curricular adaptation for developing countries is limited. Many of the studies and applications are designed for developed countries. The proposed system is designed around the Sri Lankan Education curriculum with helping tools in the child's native language to boost the child's learning trajectory.

1.3 Research Gap

While technology-mediated language interventions for autistic children have been studied with respect to user perspectives in vocabulary development, social communication support, and speech generation, there remain several critical missing elements from the literature: methods for directing language learning at dialogue level, assessing pragmatic knowledge, vocabulary-constrained personalization, tolerant speech assessment, and curriculum-aligned learning progress tracking. The following studies (Research A-E), selected from the literature review, provide an example of each limitation individually.

Comparison of Former Research

Table 2 - Comparison of former research

Application Reference	Dialogue Word Teaching	Pragmatic Context Assessment	Vocabulary Constrained Personalization	Tolerant Speech Assessment	Curriculum Aligned Progression
Research A	X	X	X	X	X
Research B	X	X	X	X	X
Research C	✓	X	X	X	X
Research D	✓	X	X	X	X
Research E	✓	X	X	X	X
Proposed System	✓	✓	✓	✓	✓

Research "A" conducted a literature review of digital technologies supporting social communication for children and adolescents with Autism. The review mapped the availability of virtual agents, mobile applications and video-modeling tools. The authors concluded that few technologies progressed beyond targeted single-skill training to functional social communication in naturalistic social settings. The review did not find a system assessing a child's understanding of the appropriate context for using a dialogue phrase or creating dialogues using words known by the child. This confirms that existing technologies provide specific dialogue skills but do not provide an assessment of a child's realistic pragmatic understanding or use vocabulary-based dialogue personalization.[1]

Research "B" reviewed technology-improved vocabulary interventions for children with ASD which included 22 studies utilizing tablet-based technology, computer-assisted technology, and robot-mediated technology. The authors also found that most studies reported positive effects on receptive and expressive vocabulary, but that most interventions were reliant on predetermined lists of words without taking children's prior knowledge into account. The authors concluded that future systems should explore "personalized and adaptive approaches" in response to individual learning trajectories. None of these studies tested the child's awareness of when, why, and how to use a word in conversation. These studies only confirmed whether or not the child could use the label to correctly identify the object. Rather, adaptive vocabulary instruction, sensitive to context, is missing and is possible using the proposed three-phase acquisition system suggested.[6]

Research "C" presented ASD-iLLM, a fine-tuned large language model for conducting dialogue-based interventions with autistic children based on Applied Behavior Analysis (ABA) principles. The system represents a significant advance by training an LLM on real clinical dialogue intervention recordings, producing responses that more closely resemble professional interventionists. However, ASD-iLLM generates dialogue from an unrestricted vocabulary, with no mechanism to constrain generated content to words a specific child has already mastered. The system is designed as a clinical intervention tool for Mandarin-

speaking children and does not target ESL contexts, curriculum-aligned vocabulary, or structured progression through dialogue complexity levels. Additionally, the system lacks any speech assessment component; interactions are text-based, meaning the child's spoken production ability is never evaluated. This demonstrates that even the most advanced LLM-based autism dialogue systems do not personalize at the individual vocabulary level or incorporate tolerant speech assessment for young, non-native speakers.[8]

Research "D" investigated how social robots can scaffold second-language learning in young children through adaptive explanation and engagement strategies. The study used a robot that adjusted its vocabulary explanations based on child responses and demonstrated measurable vocabulary gains over seven sessions. However, the robot's content was drawn from a fixed curriculum of 34 target words, with no dynamic generation of new dialogue scenarios based on what each child had individually mastered across other learning modules. The adaptation was limited to explanation strategy (how the robot taught) rather than content personalization (what the robot taught). The system also did not assess pragmatic understanding; whether a child knew the appropriate social context for a learned phrase and did not include any speech production assessment for the child's verbal output. This highlights the gap between adaptive teaching strategy and truly personalized, vocabulary-constrained content generation.[5]

Research "E" proposed a human-mediated LLM pipeline integrated into a NAO robot to deliver social communication interventions for autistic children. The system used GPT-2 and BART to generate social scenarios and questions dynamically, reducing the limitations of fixed scripted content. However, a human mediator was required at every decision point to determine which LLM-generated content was appropriate for the child, meaning the system could not autonomously adapt. The generated content was not constrained by any learner vocabulary profile, and the pipeline had no validation layer to verify that generated language matched a child's ability level. Furthermore, the system was tested only with domain experts and a child actor rather than with actual autistic children in clinical settings, leaving its real-world effectiveness unverified. No speech assessment mechanism was included; the robot delivered content verbally, but the child's spoken responses were evaluated solely by the human mediator, not by an automated system.[9]

Why Current Approaches Are Insufficient

Based on an analysis of Research A through E, five shortcomings on existing approaches are identified, establishing the need for the proposed Adaptive Dialogue Learning and Progression Tracker.

First, there are no interventions for children that teach dialogue at the level of pragmatics, that is, when and why to say a particular dialogue phrase in a particular social situation, rather than just naming objects, repeating a word, and answering questions. None of the interventions assessed a child's understanding of the same. Practical language, or knowledge

of the contexts in which to say "thank you" and other phrases, is a core deficit of ASD, but technology reviewed does not operationalize practical language through a structured and assessable learning phase.

Second, vocabulary-constrained personalization is not present. ASD-iLLM and Mishra et al. use LLMs to generate text dynamically, but do not constrain generated text to only include words that a child has confirmed learning. This can be problematic because the vocabulary generated could be new or more advanced, which could confuse or overload the student, especially autistic ESL students who are still learning the language.

Third, speech assessment for this population requires tolerance that current systems do not provide. Children with ASD exhibit atypical speech patterns including echolalia, non-standard prosody, and mispronunciation. When these children are also non-native English speakers, speech-to-text transcription errors compound further. Bosker demonstrated that fuzzy string matching can handle transcript variations in intelligibility studies, and Pham and Nguyen showed that ASR accuracy drops significantly for non-native child speakers. Yet no existing autism dialogue system integrates a multi-layered speech assessment pipeline combining empty-check, keyword matching, and fuzzy matching to accommodate these compounded inaccuracies.

Fourth, cross-module learning data is not leveraged. Existing tools operate in isolation; a vocabulary app does not inform a dialogue app about which words the child knows. Farzana et al. and Kumar et al. both developed AAC systems that function as standalone modules with no mechanism to read mastery data from other learning components. The absence of a shared learner profile means that dialogue activities cannot be automatically calibrated to each child's actual vocabulary state across all learning domains.

Fifth, curriculum alignment and element-based progression tracking are missing. Urrea et al. noted that most technology interventions use researcher-defined word lists rather than government curriculum standards. For the Sri Lankan context, where Grade 3 students are expected to produce self-introductions and descriptions following specific syllabus requirements, no reviewed system decomposes these tasks into individually trackable communicative elements or provides partial-credit assessment aligned with how special education teachers actually grade these children.

Justification of Novelty

The proposed component addresses all five identified gaps through three original research contributions that are not found in any of the reviewed literature.

Novelty 1: Three-Phase Dialogue Word Acquisition with Pragmatic Context Assessment.

Unlike existing systems that teach words through repetition or labelling alone, the proposed module introduces a structured three-phase progression with familiarization, production practice, and contextual understanding check for every dialogue word. The third phase, where the child must select the correct social scenario in which a word is used (e.g., identifying that "thank you" fits a gift-receiving scene rather than a falling-down scene), is entirely absent from all reviewed works. This operationalizes pragmatic understanding as a measurable, assessable learning objective rather than a byproduct of general social interaction, directly addressing the gap identified across Research A and B.

Novelty 2: Vocabulary-Constrained LLM Dialogue Generation with Cross-Module

Mastery Integration. While ASD-iLLM and ASD-Chat use LLMs for dialogue generation, neither constrains output to an individual child's mastered vocabulary. The proposed system reads from a shared learner profile database that stores mastery data from all four modules (concept learning, writing, pronunciation, and dialogue), then generates conversation turns using the Gemini API with an explicit instruction to use only confirmed mastered words. A post-generation validation layer checks every turn against the mastery profile before it reaches the child. This dual mechanism constrained prompting plus automated validation ensures that no unfamiliar vocabulary is ever introduced, a guarantee that no existing system provides. The approach also draws on recent advances in vocabulary-constrained text generation using LLMs for language learning contexts.

Novelty 3: Multi-Layered Tolerant Speech Assessment with Partial-Credit Element

Tracking. No reviewed system combines automated speech assessment with fuzzy matching tolerance for autistic non-native child speakers. The proposed pipeline processes each spoken response through three sequential layers: empty/no-attempt detection, exact keyword matching against expected response triggers, and fuzzy string matching (Fuzzball, threshold $\geq 80\%$) to accommodate mispronunciation and speech-to-text errors. This is grounded in Bosker's demonstration that fuzzy matching achieves reliable automated assessment of imperfect transcripts, and Pham and Nguyen's finding that ASR models require adaptation for non-native child speakers. For Level 3 sentence production tasks, the system further decomposes each syllabus requirement into individual communicative elements (name, age, location) and tracks each element across sessions using partial-credit scoring — meaning a child who says their name and age but not location is recorded as achieving two of three elements. This element-based approach aligns with how special education teachers assess these children and produce evidence-based progress reports, addressing the curriculum alignment gap identified across all five reviewed works.

The convergence of evidence from the reviewed literature provides a strong justification for the proposed system. Jiang and Cohen's systematic review of 47 studies confirmed that digital technologies for autistic children overwhelmingly target isolated skills rather than functional dialogue use, explicitly calling for tools that address "the pragmatic and social dimensions of communication". Urrea et al.'s review of 22 vocabulary learning studies found that none employed adaptive content personalization based on individual mastery profiles, recommending that "future research should explore personalized approaches". Lai et al.'s ASD-iLLM advanced LLM-based dialogue intervention but acknowledged limitations in personalization and the absence of speech-level assessment. Vogt et al. and Mishra et al. demonstrated that robots and LLMs can facilitate language learning interactions but did not achieve autonomous vocabulary-constrained content generation or automated speech assessment. Bosker and Pham and Nguyen independently provided the technical foundations, fuzzy matching and non-native ASR adaptation that the proposed speech assessment pipeline builds upon, yet neither applied these techniques within an autism dialogue learning context. Taken together, these works establish that the gap is not in any single dimension but in the integration of dialogue-level teaching, pragmatic assessment, vocabulary-constrained personalization, tolerant speech evaluation, and curriculum-aligned progression tracking within a unified system which is precisely what the proposed module aims to do.

2. OBJECTIVE

2.1 Main Objective

The primary objective of this research is to design a dynamic dialogue learning and progression tracking component that guides autistic children following the government English language curriculum from grades 1 to 3 to learn, practice and demonstrate functional English vocabulary and natural conversation flow through a structured three phase module achieving at least $\geq 75\%$ in speech assessment accuracy and $\geq 70\%$ learner task completion rate within the nine-month project timeline. By using mastery gated progression logic and mastered vocabulary constrained LLM conversation generation alongside a speech assessment pipeline with fuzzy-string matching the module aims to produce personalized dialogue activities that adapt to each child's learning level. Including non-verbal pathways and partial credit scoring ensures the equal participation of minimally verbal learners. This research aims to bridge the gap between isolated word recognition applications and real-world conversation competence, delivering a measurable improvement in conversation building skills for autistic children within special education settings.

2.2 Specific Objectives

There are five specific objectives that must be achieved to achieve the primary objective described above.

1. Implement a Three-Phase Dialogue Word Acquisition Pipeline

This focuses on creating a structured instructional sequence comprising Dialogue word and phrase introduction (image, word and audio association), Contextual Dialogue Practice (guided sentence-level exercises within themed scenarios), and a Contextual awareness check (contextual-understanding questions that confirm pragmatic grasp, not just recall). The pipeline must present each target word across all three phases before marking it as acquired, ensuring the learner demonstrates both recognition and contextual understanding. Achieving this objective enables the module to move beyond memorization toward functional vocabulary use in dialogue contexts.

2. Develop a Vocabulary Constrained LLM Dialogue Generation Engine with Validation

Here it is necessary to build a dialogue generation service that leverages a large language model (Gemini API) to produce learner level conversation prompts constrained to the learner's currently mastered vocabulary set. A post-generation validation layer will verify that every generated dialogue line contains only words within the child's known vocabulary plus the single new target word, rejecting and regenerating any output that does not adhere to these requirements. This ensures that dialogue difficulty scales precisely with each individual learner's progression rather than following a fixed curriculum.

3. Build a Multi-Layered Speech Assessment Pipeline with Fuzzy String Matching

This sub-objective targets the development of a four-stage speech evaluation system: (i) Google Cloud Speech-to-Text transcription configured for child-speech acoustic models, (ii) text normalization (lowercasing, punctuation stripping, filler-word removal), (iii) Levenshtein-distance-based fuzzy string matching against expected responses, and (iv) element-based partial-credit scoring via a spaCy Natural Language Process (NLP) microservice that awards marks for individual correctly produced keywords and phrases even when the full sentence is incomplete. This layered approach accommodates the pronunciation variability and non-standard speech patterns characteristic of young autistic ESL learners.

4. Design a Mastery Gated Progression and Non-Verbal Fallback System

The implementation flow of this sub objective is to create a progression controller that enforces four transition rules: a word advances from Phase 1 to Phase 2 only upon correct identification, from Phase 2 to Phase 3 only upon achieving $\geq 70\%$ dialogue accuracy, from Phase 3 to Mastered only upon passing the comprehension check, and any word failing three consecutive attempts triggers a guided re-teaching loop. Additionally, a non-verbal fallback pathway will be presented so that if speech input fails or the child is non-speaking, the system offers tap-to-select image-based response options, ensuring that learning and assessment continue without voice dependency. This guarantees inclusive participation regardless of verbal ability.

5. Evaluate Speech Assessment Accuracy, Task Completion, and Educational Validity

This sub-objective involves defining and applying metrics to assess the quality and effectiveness of the module across three tiers. Functional correctness will be measured through unit and integration tests covering dialogue generation, speech processing, progression logic, and API contracts. Speech assessment accuracy will target $\geq 75\%$ transcription word-accuracy, $\geq 80\%$ fuzzy-match precision, $\geq 75\%$ fuzzy-match recall, and $\geq 85\%$ partial-credit assessment instead of binary pass or fail. Educational validity

will be evaluated through a controlled pilot study (8-12 participants) measuring pre/post vocabulary gain $\geq 20\%$, task-completion rate $\geq 70\%$, average session duration 8–15 minutes, progression-unlock rate $\geq 60\%$, educator satisfaction $\geq 4.0/5.0$ via the System Usability Scale, and non-verbal pathway utilization rate. By systematically evaluating these aspects, the research ensures the module is reliable, functional, and practical for real-world special education deployment.

3. METHODOLOGY

This section outlines the research approach, technical methods and algorithms, tools and platforms, data collection procedures, ethical considerations, and validation strategy adopted in the design and development of the proposed component. All decisions reflect the final architecture, incorporating revisions made following ongoing design reviews, including the adoption of a microservice-based backend architecture for Level 3 natural language processing.

3.1 Research Approach

This research adopts a **Design Science Research (DSR)** methodology, the most appropriate approach for applied computing research that produces and evaluates a purposeful system intended to solve a defined real-world problem. DSR is particularly well-suited because the primary research output is a functional software system rather than a theoretical model or an empirical observation of an existing phenomenon.

Table 3 - The application of DSR in the research

DSR Cycle	Application for This Research
Relevance Cycle	Requirements were gathered through direct visits to a special education school in Sri Lanka, structured interviews with educators who teach autistic children, and a detailed review of the government primary English curriculum for Grades 1 to 3. This grounded the problem definition in genuine practitioner need rather than theoretical assumption, and the school became a formal research stakeholder
Design Cycle	The three-level architecture, speech assessment pipeline, adaptive mastery model, and partial credit assessment engine all emerged through iterative design informed by educator feedback, supervision, and published literature. The decision to adopt a microservice architecture for Level 3 NLP processing is itself a product of this iterative design process
Rigour Cycle	Design decisions are grounded in existing knowledge bases: peer-reviewed literature on pragmatic language intervention, adaptive educational systems, and LLM-based content generation. The validation strategy in Section 3.5 closes the rigour cycle by evaluating the artefact against defined performance metrics in a real educational context.

In addition to DSR, the research incorporates an **experimental component** at the validation stage. A structured pilot study conducted with autistic children at the partner school will evaluate system performance against predefined metrics. This combination of constructive design and experimental evaluation is consistent with applied computing research norms in educational technology for special needs populations.

3.2 Technical Methods and Algorithms

The component integrates six distinct technical methods. Each addresses a specific functional requirement of the adaptive dialogue learning system.

Three-Phase Structured Instruction (Level 1)

Level 1 teaching follows a **structured scaffolded instruction model** grounded in principles from Applied Behavior Analysis (ABA) and evidence-based practices for autistic learners. Every dialogue word from the government syllabus is taught through three mandatory sequential phases:

- **Phase 1 - Familiarization:** The child watches the word used in a contextually appropriate scene. No response is required. The system tracks how many exposures each child needs before administering a recognition check (image-tap task) as a gate before Phase 2. This ensures the child understands the meaning before being asked to produce the word.
- **Phase 2 - Production Practice:** The child attempts to speak the word or phrase. The avatar models the production with lip-sync animation first. The spoken response is captured and assessed through the speech assessment pipeline described in Section 3.2.3. A non-verbal fallback pathway is triggered if no speech attempt is detected after three prompts.
- **Phase 3 - Contextual Understanding Check:** Two images are presented side by side. The child taps the image depicting the situation in which the dialogue word would appropriately be used. This phase assesses **pragmatic language competence**; the child's understanding of when and why a word is used rather than pronunciation or recognition alone. This is the primary novel contribution of Level 1 and distinguishes this module from all other modules in the system.

Mastery-Gated Adaptive Progression

The module employs a **rule-based mastery detection algorithm** to determine when a child is ready to advance to the next word or level. It is a set of four explicitly defined rules applied at the end of every session, based on the child's recorded performance across sessions.

Each word in the learner profile exists in one of four states at any point in time:

- **NOT_STARTED** - the word has not yet been introduced to this child.
- **IN_PROGRESS** - the child has had at least one session on the word but has not yet met the mastery criteria.
- **MASTERED** - the child has met the mastery criteria across two separate sessions, and the word is added to the shared learner profile for use in Level 2 dialogue generation.

- **STRUGGLING** - the child has failed to pass three consecutive sessions for this word. The word is deferred, and a simpler word is introduced instead.

A session is considered a pass when the child achieves a Phase 2 production score of 2 or above (at least a partial sentence) and answers the Phase 3 contextual check correctly. A score of 1 (single word only) or an incorrect contextual check does not count as a pass toward mastery. The four state transition rules are:

- **Rule 1 - Mastery promotion:** If the child has passed the session for a word on two or more occasions on different calendar days, the word transitions from IN_PROGRESS to MASTERED. Requiring different days prevents a child from achieving mastery through repeated attempts within a single sitting, which would not demonstrate durable learning.
- **Rule 2 - Struggling detection and deferral:** If the child has failed three consecutive sessions for the same word, the word transitions to STRUGGLING and is deferred. The system identifies the next available word with a lower difficulty rating in the same or a related category and introduces it instead. This prevents a child from being repeatedly exposed to a word that is currently beyond their capacity, which can produce avoidance behaviors in autistic learners.
- **Rule 3 - Return from deferral:** Once all words with a lower difficulty rating in the same category are mastered, the deferred word is re-introduced with its fail streak reset to zero and its status returned to IN_PROGRESS. Phase 1 is restarted with new contextual images to provide a fresh exposure before assessment is attempted again.
- **Rule 4 - Pacing adaptation:** The system records the number of sessions each child requires to master each word. If a child achieves mastery in one session (maximum score on first attempt), Phase 1 familiarization for subsequent words is shortened from four exposures to two. If a child requires three or more sessions, Phase 1 familiarization for subsequent words is extended, and additional contextual image pairs are shown before Phase 2 is attempted. This adapts the teaching pace to the individual child's demonstrated learning rate without changing the mastery criteria themselves.

An asset library will be created with the vocabulary covered in the Grade 1 to 3 Government curriculum. Words in the word library are grouped into three difficulty levels. Difficulty 1 covers the simplest words: hello, goodbye, yes, no, okay, please, and thank you. Difficulty 2 covers sorry, help, good morning, and good afternoon. Difficulty 3 covers multi-word classroom phrases such as excuse me, may I come in, and may I sit. The algorithm always introduces difficulty 1 words first. Difficulty 2 words are gated behind difficulty 1 mastery within the same category, and difficulty 3 words are gated behind difficulty 2.

Speech Assessment Pipeline

The speech assessment pipeline converts a child's spoken response into a structured score that determines the feedback delivered. The pipeline operates in four sequential stages:

Table 4 - Speech assessment pipeline breakdown

Stage	Description
Stage 1 - Audio Capture	The child's spoken response is captured via the device microphone using Expo AV within the React Native application. The audio is recorded as an M4A file and transmitted to the Node.js backend.
Stage 2 – Speech to Text Transcription	The audio file is forwarded to the Google Cloud Speech-to-Text API. The en-GB language configuration is used as it provides better recognition accuracy for South Asian English accents than en-US. The API returns a lowercase text transcript of the child's spoken response.
Stage 3 - Transcript Analysis	The transcript is analyzed through three sequential layers: (1) an empty check that detects silence or filler-only responses and triggers the non-verbal fallback flow; (2) exact keyword matching against predefined trigger lists for each response level: full sentence, partial sentence, and single word; and (3) fuzzy string matching using the fuzzball library for cases where exact matching fails due to mispronunciation or transcription error.
Stage 4 - Score Assignment	A score of 0 (no attempt), 1 (single word), 2 (partial sentence), or 3 (full target sentence) is assigned. The corresponding feedback message is selected, the avatar delivers it with lip-sync animation, and the result is persisted to the learner profile in the PostgreSQL database.

Fuzzy String Matching. A key design decision is the use of fuzzy string matching as a fallback when exact keyword matching fails. Autistic children learning English as a second language are likely to produce phonologically approximate rather than exact outputs. Speech-to-text systems may further distort these into non-standard transcriptions, for example, a child saying "thank you" may produce a transcript of "tank you" or "fank you", both of which would fail exact matching and penalize a genuine communicative attempt. The fuzzball library implements the Levenshtein distance-based partial ratio algorithm, returning a similarity score between 0 and 100. A threshold of **80%** has been selected as the initial acceptance boundary, based on the principle that scores at or above this level reliably indicate communicative intent while filtering genuinely incorrect responses. This threshold will be empirically calibrated during pilot testing, and the calibration data will be reported in the final research submission.

Non-Verbal Fallback Pathway

The component incorporates a structured **non-verbal fallback pathway** for children who do not produce a spoken response during Phase 2 of Level 1. The pathway is triggered after three prompted attempts with a 20 to 30 second response window between each prompt. The fallback pathway presents a Word-to-Scene Matching activity in which three images are displayed, and the child taps the image showing when the word is used. This activity:

- Assesses recognition and associative understanding without requiring spoken output.
- Uses three options rather than two to reduce the probability of correct guessing.
- Records the result as **recognized but not yet verbally produced** in the learner profile, preserving the distinction between receptive and expressive language achievement.
- Provides educators with meaningful information about a child's receptive knowledge even in the absence of verbal production, surfaced as a separate achievement category in the teacher report.

Vocabulary-Constrained LLM Turn Generation (Level 2)

Level 2 uses a **vocabulary-constrained large language model prompting strategy** to dynamically generate personalized conversation turns at session time. The system constructs a structured prompt from each child's live mastery profile and submits it to the Google Gemini API, constraining the model to generate dialogue using only vocabulary the child has confirmed mastery of across all four modules.

The prompt explicitly instructs the model to:

- Use only vocabulary items present in the child's confirmed mastery profile.
- Produce output in a strict JSON schema matching the turn library data structure.
- Include exactly one greeting turn, one farewell turn, and four to six content turns sequenced from difficulty level 1 to level 3.
- Include at least one politeness turn if the relevant dialogue words are in the mastery profile.

A **validation layer** sits between the LLM response and the session delivery pipeline. It checks that all generated turns reference only mastered vocabulary, that required JSON fields are present, and that all three response levels are defined. Turns that fail validation are discarded and replaced with entries from a manually authored seed library of 20 to 30 turns, which also serves as a fallback for offline scenarios. This validation layer ensures that no unfamiliar vocabulary ever reaches a child regardless of LLM output quality.

Element-Based Partial Credit Assessment with spaCy Microservice (Level 3)

Level 3 assesses spoken production on Year 3 syllabus topics through an **element-based partial credit model**. Each topic of self-introduction, describing a friend, describing a pet is decomposed into its smallest communicable units, referred to as elements. Achievement is assessed and recorded at the element level rather than as a binary pass or fail at the task level.

Natural language processing for element detection is handled by a dedicated **Python microservice** running the **spaCy** library. This service runs independently from the Node.js main backend and is called via internal HTTP request when a Level 3 assessment is triggered. The microservice architecture was chosen because spaCy is Python-native and provides significantly more reliable named entity recognition than any available Node.js NLP library for this use case. A fallback is built into Node.js: if the Python service is unavailable, the system falls back to keyword-only matching without spaCy entity detection, so Level 3 remains functional.

The element detection pipeline within the Python microservice operates through three sequential layers:

- **Layer 1 - Keyword matching:** The transcript is checked for the presence of keywords and phrases associated with each element. For example, detection of any number word combined with "years" or "old" indicates that the age element was attempted.
- **Layer 2 - spaCy named entity recognition:** For elements requiring entity identification such as whether the child mentioned their own name (PERSON entity) or a location (GPE or LOC entity) spaCy's pre-trained entity recognizer is applied to the transcript.
- **Layer 3 - Child profile validation:** For name and location elements, the detected entity is cross-checked against the child's stored profile data. This prevents a child being credited for mentioning someone else's name when their own was expected.

Grammatical correctness is not required for an element to be marked as achieved. A child who produces "hair long" in response to a prompt about their friend's appearance is credited with the appearance element, the system assesses communicative intent, not grammatical form. Element mastery requires consistent achievement across a minimum of **two separate sessions**.

The teacher reports will be generated using a **translation layer**, converting the child's raw interaction data collected by the system and translating them into explainable reports for the teacher.

Prior to the spoken production assessment activities, Level 3 includes a contrastive sentence structure instruction phase designed to address first language interference in English sentence construction. Sri Lankan children whose first language is Sinhala, operate with a Subject-Object-Verb (SOV) sentence order, whereas English follows Subject-Verb-Object (SVO)

order. To address this, the system introduces a brief contrastive instruction sequence before the child attempts sentence production. The sequence first presents the target sentence in Sinhala word order, so the child sees a familiar structure, then contrasts it directly with the English equivalent, visually highlighting how the verb position changes between the two languages. Each sentence part is color-coded across both languages using three consistent labels: WHO (Subject), ACTION (Verb), and WHAT/WHERE (Object or Complement), with the Sinhala label displayed beneath each English label to reduce cognitive load. This color-coded scaffolding constitutes enhanced input drawing the child's attention to sentence structure implicitly without requiring explicit explanation. Following the contrastive demonstration, the child completes a word ordering drag-and-drop activity in which scrambled word tiles are placed into the correct WHO, ACTION, and WHAT slots, consolidating the structural pattern through active engagement. Sinhala is used exclusively for instructional labels and slot headings within this activity but all sentence content, word tiles, and audio remain in English ensuring that native language support reduces task load without displacing the target language.

3.3 Tools and Platforms with Justification

The following tools and platforms were selected based on four criteria: suitability for the specific technical function, quality of support for the target use case (children's speech, mobile deployment, special education), availability within a research prototype budget, and coherence with the overall four-module system architecture.

Table 5 - Tools and Platforms to develop the proposed module

Tool / Platform	Purpose	Justification
React Native with Expo	Cross-platform mobile frontend	Enables a single codebase for both Android and iOS. Expo AV provides reliable microphone access and audio playback.
Node.js with Express.js	Main backend API server	Shares JavaScript syntax with the frontend, reducing context-switching. Its asynchronous non-blocking architecture handles the I/O-heavy workload well, given multiple concurrent calls to Google Speech to Text and the Gemini API per session.
Python with Flask	spaCy microservice for	spaCy is Python native and provides the most reliable named entity

	Level 3 element detection	recognition available. A dedicated Flask microservice isolates this Python dependency cleanly.
spaCy	Named entity recognition within the Python microservice	Identifies PERSON, GPE, and LOC entities within speech transcripts, enabling Level 3 element detection without requiring exact string matches.
PostgreSQL	Relational database for learner profiles and content	The learner profile data has an inherent relational structure. PostgreSQL provides ACID compliance ensuring session results are never partially written, and supports complex joined queries for mastery tracking and teacher report generation.
Google Cloud Speech to Text	Speech transcription	Provides the most reliable recognition for children's voices and non-native English accents. The en-GB language model performs better with South Asian English phonology than en-US.
Google Gemini API	LLM-based turn generation (Level 2)	Selected for consistency with the concept learning module.
fuzzball (Node.js)	Fuzzy string matching for speech assessment	A JavaScript port of FuzzyWuzzy implementing the Levenshtein partial ratio algorithm. Provides equivalent functionality within the Node.js backend.
ElevenLabs	Avatar voice generation	Produces natural, consistent voice output more suitable for child facing educational content. Pre-generation eliminates runtime TTS latency and supports offline operations.
Rhubarb Lip Sync	Avatar lip-sync animation data	An open-source tool that analyses audio files and outputs mouth shape timing data in JSON format. All avatar audio is processed offline, eliminating runtime lip-sync computation.
Lottie React Native	2D avatar rendering	A Lottie-based 2D animated avatar is proposed as the animations are lightweight, render consistently on mobile, and are appropriate for child-facing applications.

Docker	Containerized deployment	Ensures that both the Node.js backend and Python microservice run consistently across environments. Docker Compose manages both services together, and the internal Docker network allows private communication without external exposure.
---------------	--------------------------	--

3.4 Data Collection and Ethical Considerations

Data Collected by the System

The component collects two categories of data during normal operation:

- **Interaction data:** Turn-by-turn session results including the speech transcript, response level achieved, score, match type (exact or fuzzy), and timestamp. This data drives mastery tracking, adaptive session generation, and teacher report production.
- **Learner profile data:** Mastery status per word and per element, session history, pacing metrics, and struggle flags. This data persists across sessions and is shared in read-only form with the other three modules through the shared learner profile database table.

Pilot Study Data Collection

A pilot study at the partner school will evaluate system performance. Data collected during the pilot will include:

- Session completion rates per child per level.
- Turn-level response scores tracked across sessions.
- Speech assessment accuracy measured against manually reviewed ground truth labels.
- Educator ratings of teacher report usefulness via a structured Likert-scale questionnaire.
- Observational notes from educators on child engagement and distress indicators.

No video recordings of children will be made during the pilot. All data will be anonymized before analysis, with children identified only by assigned participant codes.

Ethical Considerations

The following ethical considerations have been addressed in the design of the system and research process:

- **Informed consent:** Written informed consent is taken from the partner special education school. Consent is voluntary and may be withdrawn at any time without consequence.
- **Child assent:** Where possible, age-appropriate and communication-appropriate assent will also be sought from the child participant, in recognition of the child's right to participate in decisions about their own involvement.
- **Data minimization:** The system collects only the minimum data necessary for its educational and research functions. Personally identifiable information is stored separately from interaction data.
- **Data security:** All data is stored in a password-protected database with access restricted to the researcher and supervising academic. No data will be shared with third parties beyond anonymized aggregate findings in the research submission.
- **Vulnerability of participants:** Autistic children constitute a vulnerable population. The system has been designed with specific safeguards: the non-verbal fallback pathway, the absence of failure messages, positive reinforcement at every response level, and a maximum attempt limit before redirection to prevent distress. Educators will be present during all pilot sessions and may terminate a session at any point.
- **Right to withdraw:** If a child shows signs of distress during a session, the session will be terminated immediately by the supervising educator. No penalty attaches to non-completion.
- **Benefit to participants:** The research is designed to produce a tool that directly benefits the children who participate in its evaluation. Pilot participants will have continued access to the system following the study period if the school and families wish.

3.5 Validation Strategy and Performance Metrics

Validation of the component is conducted at three levels: **functional correctness** (does the system behave as designed?), **technical accuracy** (does the speech assessment pipeline and the spaCy microservice produce reliable results?), and **educational validity** (does the system produce measurable learning outcomes and useful teacher evidence?). Each level uses distinct methods and metrics.

Functional Correctness

Unit and integration testing will be performed on all backend components prior to the pilot study:

Table 6 - Functional testing criteria for the proposed module

Test Area	Pass Criterion
Turn generation	The conversation builder is called with 20 mock learner profiles of varying vocabulary breadth. All generated sessions must contain only turns whose mastery requirements are satisfied by the mock profile, and every session must contain exactly one greeting and one farewell turn.
LLM validation layer	80 LLM-generated turn sets are submitted to the validation layer, including intentionally malformed sets. All turns referencing non-mastered vocabulary must be correctly rejected; all structurally incomplete turns must be correctly discarded.
Mastery gating	Session result sequences are injected into the mastery tracking algorithm covering all edge cases. Words must be promoted to mastered status only after achievement across two non-consecutive sessions.
Non-verbal fallback	Audio inputs simulating silence and filler sounds are submitted to the speech pipeline. All empty or filler-only inputs must correctly trigger the non-verbal fallback pathway without score assignment.
Python microservice connectivity	The Node.js backend is tested with the Python microservice available and unavailable. When available, element detection results must be correctly returned and saved. When unavailable, the Node.js fallback keyword matching must activate without crashing the session.
spaCy entity detection	30 test transcripts per Level 3 topics, covering correct responses, partial responses, and irrelevant responses are submitted to the Python microservice. Element detection results are manually reviewed for correctness. Target: ≥85% accuracy across all element types.
Database integrity	All session endpoints are tested with valid and invalid inputs. No partial writes must occur; all foreign key constraints must be enforced; session results must be correctly associated with the correct child, session, and turn.

Speech Assessment Accuracy

The accuracy of the speech assessment pipeline will be evaluated against a **manually labelled reference set** of 30 to 50 audio recordings collected from children at the partner school during informal pre-pilot sessions. Each clip will be independently labelled by the researcher and a supervising educator with the response level it represents.

Table 7 - Speech assessment criteria for the proposed module

Metric	Definition and Target
Response Level Classification Accuracy	Proportion of audio clips for which the pipeline assigns the same response level as the human ground truth. Target: ≥80% across all response levels.
False Negative Rate (missed attempts)	Proportion of genuine speech attempts incorrectly classified as no attempt. This is the highest priority error class as it causes a child's communicative effort to go unrecognized. Target: below 5% .
False Acceptance Rate (wrong level)	Proportion of single-word responses incorrectly classified as full sentence. Target: below 10% .
Fuzzy Match Threshold Calibration	Following initial testing, the 80% fuzzy match threshold will be adjusted based on observed error rates. The final calibrated threshold and supporting data will be reported in the methodology of the final submission.
Transcription Coverage	Proportion of audio clips for which Google Speech-to-Text returns a non-empty transcript. Target: ≥85% , given the expected speech clarity profile of the participant population.

Educational Validity

Educational validity is assessed through the structured pilot study with autistic children at the partner school, conducted under educator supervision across a minimum of four sessions per child.

Table 8 - Educational Validity criteria for the proposed module

Metric	Definition, Measurement Method, and Target
Session Completion Rate	Proportion of started sessions completed without educator intervention to terminate due to distress. Target: ≥75% ,

	recognizing that some non-completion reflects child preference rather than system failure.
Within-Session Progression	Mean response level score across turns within a session, tracked across sessions per child. A positive trend indicates that learning is occurring. Target: positive trend observable in ≥60% of children across four sessions.
Mastery Achievement Rate	Number of Level 1 words achieving mastered status per child across the pilot period. No fixed target; this metric provides descriptive data on learning pace across participants.
Teacher Report Usefulness	Educator rating of the auto-generated teacher report on a five-point Likert scale across four dimensions: accuracy of achievement data, clarity of presentation, usefulness for grading, and usefulness for lesson planning. Target: mean rating of ≥3.5 across all dimensions.
Partial Credit Alignment (Level 3)	Educator agreement with element-level scores generated by the system compared with their own observational assessment. Target: agreement on element achievement status in ≥80% of assessed elements.
System Availability	Proportion of session attempts that are complete without a system error in either the Node.js backend or the Python microservice. Target: ≥95% .

In regard to the sample size, the partner school has a limited population of autistic children at the appropriate grade level. The pilot study is expected to involve between 5 and 10 participants depending on consent rates and availability. This sample size is not sufficient for statistically generalizable results, and no such claims will be made in the research submission. The purpose of the pilot is formative evaluation of the artefact - demonstrating that the system functions as designed, that the speech pipeline produces reliable results in the target population, and that educators find the outputs useful. This is consistent with DSR evaluation norms, which require demonstrated utility in the target context rather than population-level generalizability.

4. HIGH-LEVEL SYSTEM ARCHITECTURE

4.1 Overview of proposed solution and its components

The proposed system is a **mobile-first adaptive English learning application** for autistic children studying in Sri Lankan special education schools, built around the government primary school English curriculum for Grades 1 to 3. The full system is a collaborative research project comprising four independently developed modules that share a common learner profile database. Each module addresses one dimension of English language acquisition:

Table 9 - Proposed components of the system

Module	Function & Technology
GNN-Enhanced Adaptive Concept Learning Module	Teaches vocabulary through image recognition and activities that dynamically adapt to the learner's interactions. Provides the mastered concept words that Level 2 of this module uses for dialogue construction.
Adaptive Dialogue Learning and Progression Tracker (this module)	Teaches dialogue words, builds personalized spoken conversations, and assesses sentence production. Three levels: word acquisition, dialogue construction, sentence production.
Motor-Informed Adaptive Handwriting Learning Module	Assesses motor readiness and teaches letter formation through adaptive tracing activities and personalized to each learner's abilities.
Adaptive English Pronunciation Support Module	Teaches phoneme production from the alphabet through to short words using avatar lip-sync and pronunciation comparison scoring.

4.2 System Diagram of Components

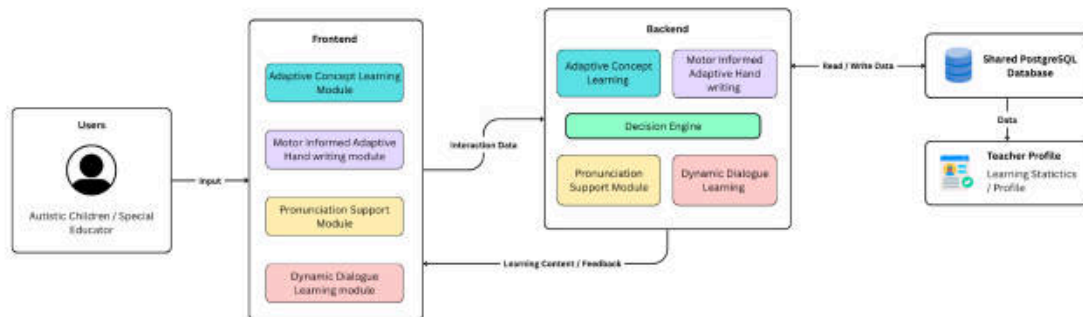


Figure 1 - High level architecture overview of the system

The above figure illustrates the full system for the proposed system. The data flow is shown where the user inputs in each separate module are taken as interaction data through the front end of the system, and the interaction data is passed to the backend of the specific module. The backend Logic of each separate module reads and writes to the shared database, and the output to the interaction data is processed through the backend and displayed in the frontend as Learning content, activities or feedback to the user. Each teacher profile will include all the children's profiles of the students assigned to the specific teacher, with the teacher's own profile details, and the student's learning progress with visuals and statistics according to each component.

4.3 Proposed Module – Component Interactions & Data Flow

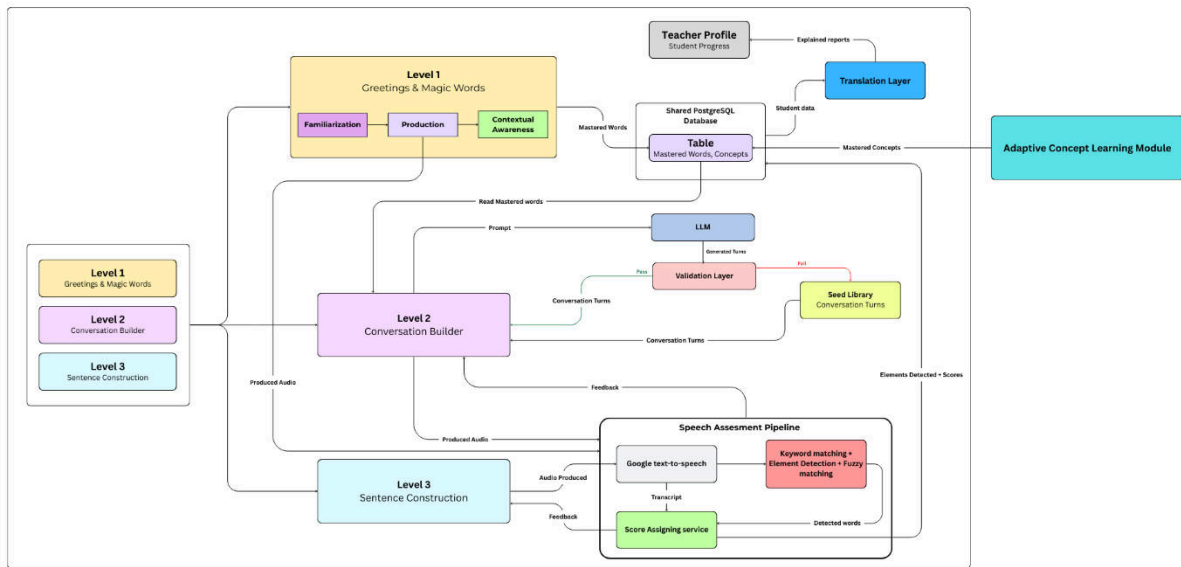


Figure 2 - High Level architecture overview of the proposed component

Figure 2 illustrates the internal architecture of the proposed module. The module consists of five internal subsystems: the three teaching levels, the speech assessment pipeline, the LLM turn generation with the seed library, and the database tables, teacher profile and translation layer.

Level 1 - Dialogue word acquisition

This level is independent of the other modules. It teaches words and phrases used in conversations through three tiers, such as Familiarization, Production, and Contextual Awareness. A non-verbal pathway is configured for children who do not produce spoken audio during the Production phase. When the child masters a word in Level 1, it will be recorded in a table in the shared database with the mastered concepts from the Concept Learning module.

Level 2 - Personalized Dialogue Construction

The level 2 will take the words that the child has mastered from the shared mastery table, and the session builder will call on the Gemini API prompt to generate conversation turns constrained by the vocabulary from the child's mastered words. The generated turns are validated by the validation layer and if more than 4 turns violate the constraints given, it will fall back and generate conversation turns using the manually created seed library. The seed library will also be used in case of an API failure to generate the conversation turns to ensure a smooth learning experience. Here the integration happens with the component Concept learning at the database level as the child's mastered concepts are also added alongside the

Dialogue Learning module's Level 1 mastered words and phrases in the mastery profile table and the mastered components are used to build personalized conversations for the student according to each student's specific knowledge level.

Level 3 - Sentence Production and Partial Credit Assessment

Level 3 dives into sentence production focusing on the topics required by the Grade 3 syllabus curriculum as self-introduction, describing a friend, and describing a pet. The child is first shown a video to help introduce the context and structure then continues in flow until the child is taught how to attempt independent sentence production. Assessment of the sentences is element based: each task is broken down into its smallest communicable units and the python spaCy microservice detects which elements the child achieved when speaking the sentences. Each achieved element is recorded, and the teacher reports are generated with these as proof.

The speech assessment pipeline is common for each of the 3 levels. The child's speech is recorded and sent to the backend where it is forwarded to Good Cloud Speech to text and receives a transcript of the audio. The transcript is analyzed in three levels: an empty check to see whether the child has spoken or not, exact keyword matching; in Level 1 to check whether the word has been exactly pronounced, in Level 2 whether the keywords are an exact match to the turn's predefined sentence list, and the last level is fuzzy matching using fuzzball in Node with an 80% similarity threshold. A score of 0 to 3 is assigned. 0 being not attempted, 1 and 2 being partially attempted and 3 being fully successful in the expected sentence production.

The teacher profile will be given explainable reports and child progression metrics by translating the raw metrics collected through the system using a translation layer.

4.4 Scalability and Security Concerns

Scalability

The proposed system is designed to serve a small pilot population in the special education school. The given considerations are for architecture decisions if the system is deployed for a wider population.

- The Node.js backend is stateless, and all session states are in the PostgreSQL database, not in the server memory. Therefore, additional Node.js instances can be deployed without architectural changes if user volume increases.

- As the python spaCy microservice is in a separate container, if the level 3 assessment load increases, additional instances can be created independently without affecting the backend.
- PostgreSQL can be scaled vertically (larger server) or horizontally (read replicas for reporting queries) without changes to the application code.
- A limitation is created for scalability since the Gemini API has per-minute rate limits. If adopted at a wider scale with a session queue and a larger seed library (expanded from 30 to 100+ turns), this will reduce Gemini dependency.

Security

Since the system handles data from vulnerable child participants, security is a primary concern when designing this system.

- All external API keys are stored in .env files and are never committed to version control. During deployment these would be managed using a secrets management service.
- Child names are only stored in the student registration table, and all the other tables refer to the child with a system generated ID to reduce identification risk in case of a data breach.
- Role based access control ensures that the accessible content in the system is constrained for each user. The student session data is not accessible to unauthorized requests.

5. USER REQUIREMENTS

5.1 Stakeholder identification

The proposed system involves three specific stakeholders. They are categorized as:

- (1) Principal – The special education school’s principal will have the role of **system administrator** within the system.
- (2) Special Education Teacher – Will be delegated the role of **proxy user** as they are the primary operational users who manage and navigate the system on behalf of the child during learning sessions.
- (3) Autistic Child – The **primary end user** of the system is the autistic children as they are the one’s following the learning content and attempting the given activities.

5.2 Functional Requirements

1. The system shall allow the system administrator to register and create individual children and teacher profiles in the system and manage and edit them when necessary.
2. The system shall allow the system administrator to assign students to a specific registered teacher in the system.
3. The system shall follow only the words and content from the Sri Lankan English language curriculum within grades one to three in its teaching content and activities.
4. The system shall follow a three-tier instruction method of Familiarization, Production, and Contextual Awareness when teaching Level 1 Dialogue words.
5. The system shall constrain the LLM to generate conversation turns using only the words the child has already mastered from the other modules.
6. The system shall allow the primary user to access the learning content through clear interfaces and large touch targets.
7. The system shall allow the child to speak during production phases by tapping the microphone icon, and the audio shall be captured for processing.
8. The system shall provide the child with clear child-friendly feedback with avatar animations and simple written phrases.
9. The system shall adapt to the difficulty of the instruction of words and phrases based on the child’s previous performance.
10. The system shall provide activities and a learning path for non-verbal children if they do not attempt to speak after several prompts.
11. The system shall allow the special education teacher to log in and view the students assigned to them.

12. The system shall allow the special education teacher to access the student's learner profile to check the student's learning progress and mastered concepts, words, and phrases across all the modules.
13. The system shall provide the special education teacher with a detailed view of the student's attempts, responses for learning content, and understanding results with dates for every activity attempted.
14. The system shall give a summary of progress at each level in the module with visual indicators.
15. The system shall allow the special education teacher to adjust session frequency and daily usage for each student.
16. The system shall generate explained teacher-friendly printable and downloadable reports aligned with the learning requirements for each grade in the government syllabus.
17. The system shall allow the principal to view high-level reports of each student.
18. The system shall allow the principal to monitor the system usage of the teachers and students.
19. The system shall present a contrastive sentence structure instruction phase in Level 3 that visually demonstrates the difference between Sinhala SOV and English SVO word order before the child attempts sentence production activities.

5.3 Non-Functional Requirements

1. Security
 - All student data shall be stored in a password protected database with access only to authorized individuals with personal data stored separately from interaction data.
2. Performance
 - Child feedback for the speech assessment must be given in under 5 seconds.
 - The turn-generation using the LLM should be completed within 10 seconds from the start of the session, and the seed library must be available if the response time exceeds, or the network is unavailable.
3. Usability
 - The interfaces shown to the child must use simple, high contrast large visuals.
 - The teacher reports must be user friendly with clear child progression data that a special educator without a technical background will be able to understand.
4. Reliability
 - If the Gemini API is unavailable, the Level 2 sessions will be created using the manually created seed library for an uninterrupted learning flow.

5. Scalability

- The system must support concurrent sessions for up to 20 children without performance issues.

6. Accessibility

- The system should be able to be used in both Android and iOS devices to accommodate various device types that may be used by the children.

6. COMMERCIALIZATION PLAN

6.1 Target market and customer persona

The primary target market is government and private schools in Sri Lanka that operate dedicated special education units or inclusive classrooms with autistic children learning English as a second language. The main customer persona is the school principal or special education coordinator responsible for resource allocation and academic outcomes in these units. A secondary persona is the special education teacher who uses the platform daily to plan lessons, run dialogue activities, and monitor learner progress, while parents benefit indirectly through improved communication and independence in their children.

6.2 Value proposition

The proposed system offers an end to end, curriculum aligned dialogue learning platform that moves autistic learners from isolated vocabulary drills to functional conversation, something traditional AAC tools and generic language apps do not prioritize. It reduces teacher preparation time by generating personalized activities automatically, while providing objective speech assessment data and progression reports that can be used for Individualized Education Plans (IEPs) and school reporting. For institutions, the platform delivers predictable subscription pricing, centralized administration, and evidence of learning gains that can support funding and accountability requirements.

6.3 Revenue model

The system will operate on a Software as a Service subscription model, with an annual license fee charged per institution rather than per student. This structure will make budgeting predictable for schools operating under fixed government allocations.

Three subscription tiers are proposed, as outlined in the table below:

Table 10 - Revenue model tiers

Tier	Student Capacity	Target Segment
Tier 1	Up to 30 students	Small special education units
Tier 2	30 – 120 students	Mid-scale schools and larger special education units
Tier 3	120+ students or multi-campus institutions	Large institutions with expanded administrative needs

In addition to the paid tiers, a freemium pilot license could be made available, to allow schools to evaluate the platform before committing to a subscription, which aligns with how procurement decisions are typically made in Sri Lankan public education, where trial evidence is often required prior to budget approval.

6.4 Cost estimation

Initial costs include system development (engineering time for dialogue engine, speech pipeline, and teacher dashboards), infrastructure (cloud hosting, storage, and API usage for LLM and speech services), and UX design tailored for autistic learners. Recurring operational costs cover cloud servers and databases, third-party API fees, security monitoring and backups, customer support, and periodic content updates aligned with curriculum revisions. Marketing and onboarding costs are expected to remain moderate, as adoption will primarily be driven through pilot partnerships with special-education units, professional networks, and ministry-level referrals rather than mass advertising.

6.5 Pricing strategy

The platform will follow a Software as a Service institutional licensing model, using the three subscription tiers defined in the revenue model so that schools only pay according to student capacity. This per institution model fits the fixed budget nature of Sri Lankan public schools and simplifies procurement compared with per student pricing. A time limited freemium or pilot license will be offered to allow schools to trial the system with a small group of learners before committing to a paid tier, reducing perceived risk for decision makers and accelerating adoption. For low resource rural schools, concessional or grant supported pricing can be explored in collaboration with NGOs and government programs focused on inclusive education.

6.6 Competitive advantage

Compared with generic ESL apps and existing AAC tools, the proposed system uniquely combines curriculum aligned content, mastery gated progression, speech-based assessment with fuzzy matching, and non-verbal participation options designed specifically for autistic learners. The dialogue engine uses mastered vocabulary as a constraint, ensuring that generated conversations remain learner appropriate and measurable rather than open ended and unpredictable. Integrated analytics, IEP-friendly reporting, and multi-stakeholder support (principal dashboards, teacher tools, child-friendly interface) further differentiate the product as a school-ready solution rather than a standalone consumer app.

6.7 Intellectual property considerations

The core intellectual property resides in the system architecture, dialogue progression logic, assessment rules, and curated content set that operationalizes the national English curriculum for autistic learners. Source code, custom datasets, and UI designs will be protected as proprietary assets; access will be granted only through the hosted SaaS platform and not by source distribution. Third party services such as LLM and speech to text APIs will be used under their respective licenses, with clear separation between proprietary business logic and commodity infrastructure. Where appropriate, trademarks can be registered for the product name and logo, while any reusable research artefacts (e.g., anonymized datasets or evaluation instruments) can be selectively open-sourced without exposing commercially sensitive components.

7. BUDGET AND JUSTIFICATION

The below given figure breaks down the budget for the proposed project with the justification.

Table 11 - Budget and justifications

Item	Cost (Total LKR)	Justification
Monthly travel to special education school	6000	Transport to the partner special education school for user testing, observations, and teacher feedback sessions.
Gemini API (dialogue generation)	4,800	Pricing for Gemini Flash / Flash-Lite tier model is roughly 0.075-0.10 USD per 1M input tokens and 0.30-0.40 USD per 1M output tokens, with a free tier.
Google Cloud Speech-to-Text	4,500	For the transcription of audio tasks, Standard Google pricing is around 0.024 USD per minute with some free quota
Development software & tools	5,000	Paid IDE extensions, design tools, or utilities not covered by free student licenses.
Contingency	5,000	Buffer for unexpected extra API usage, additional school visits, or minor equipment/software needs.
Total Cost	25,300	

8. WORK BREAKDOWN STRUCTURE (WBS)

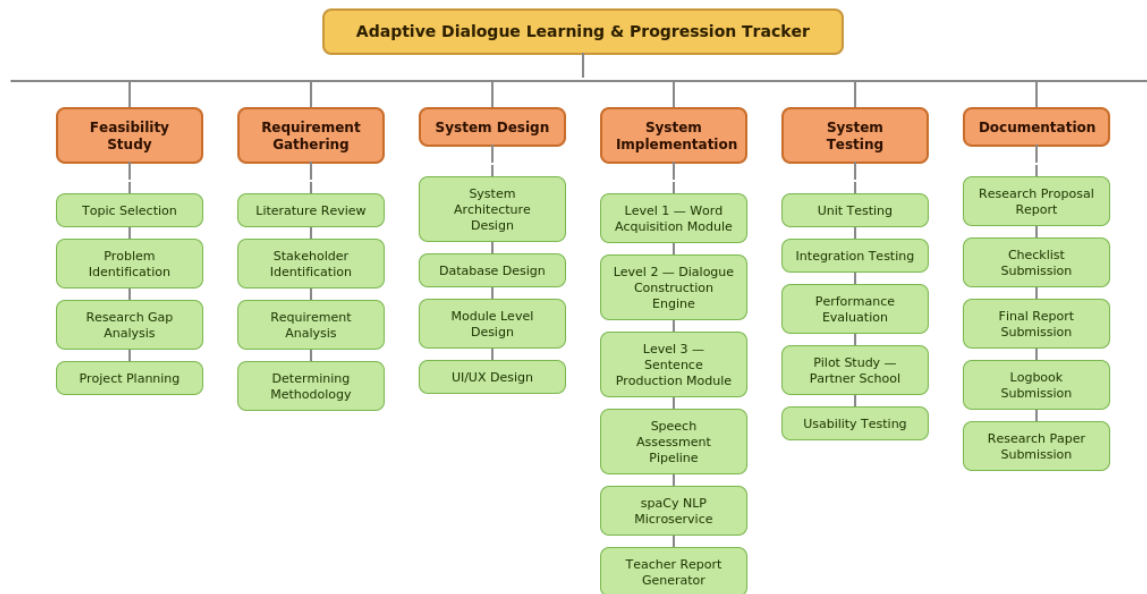


Figure 3 - WBD structure

9. GANTT CHART

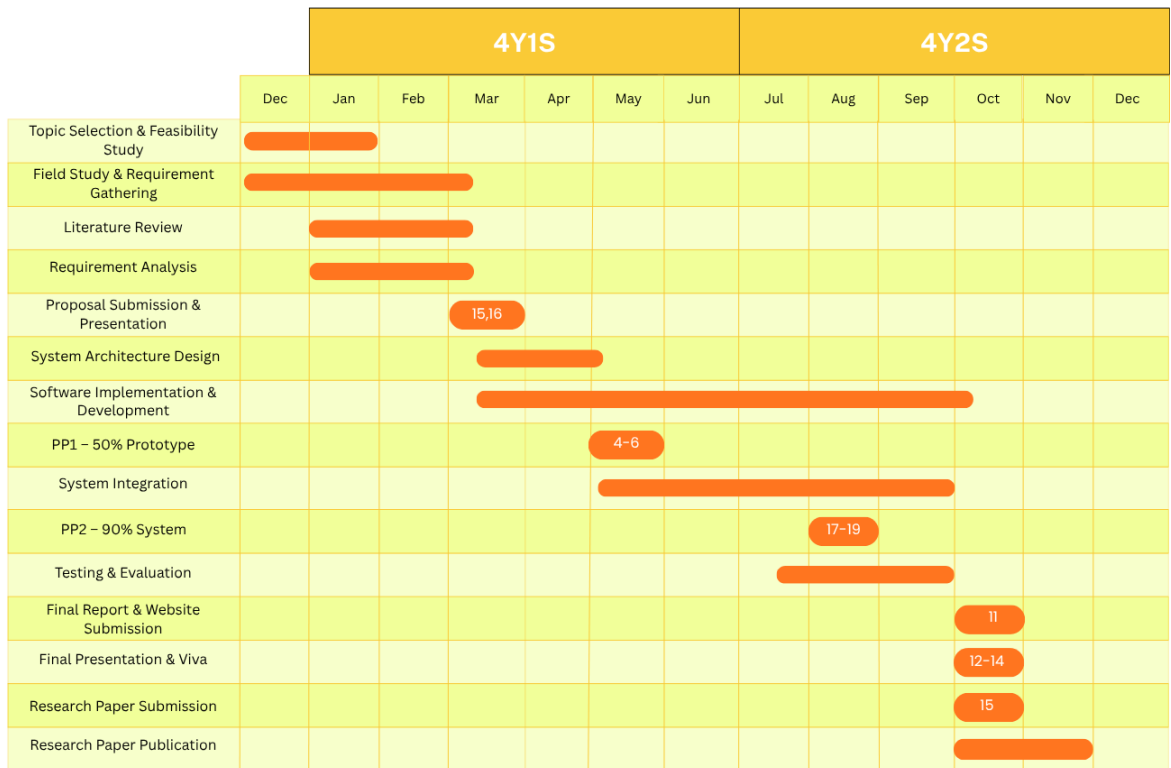


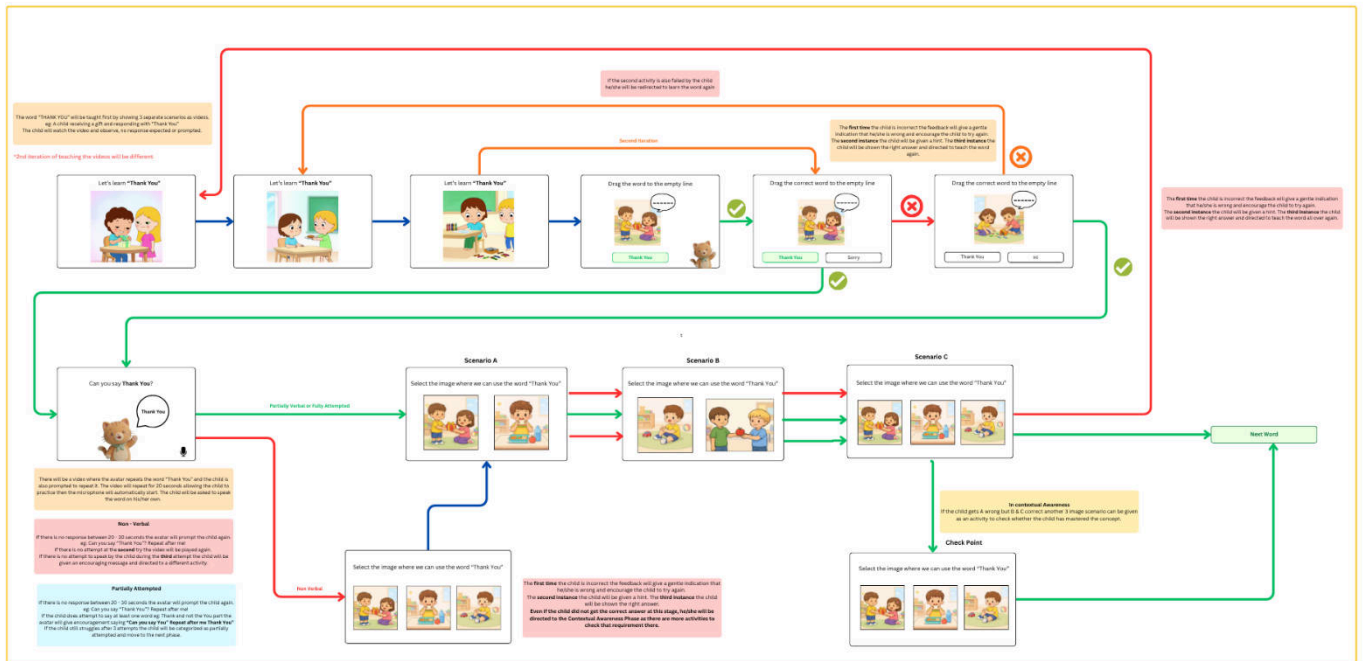
Figure 4 - Gantt Chart

10.REFERENCES LIST

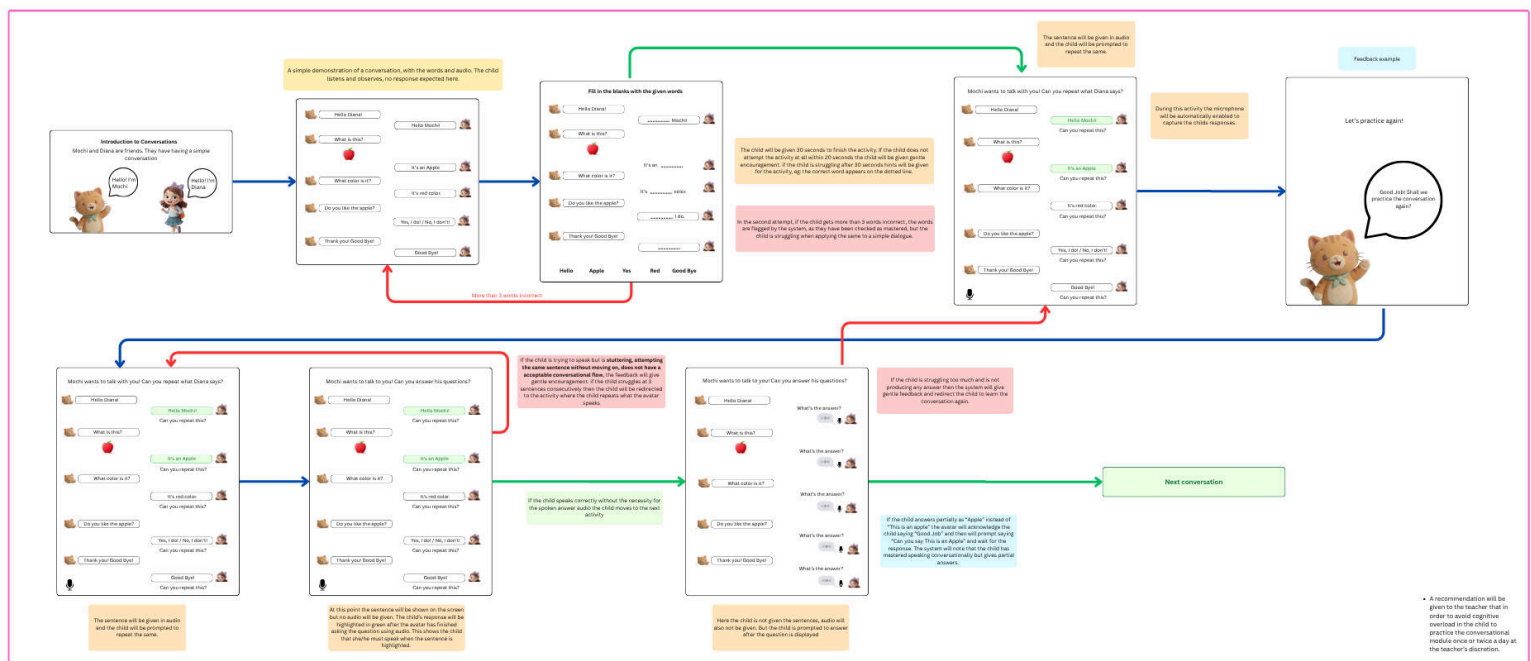
- [1] X. Jiang and S. R. Cohen, “Supporting autistic children and adolescents’ social communication skills through digital technologies: A systematic literature review,” *Research in Autism*, vol. 130, 2026, Art. no. 202750, doi: 10.1016/j.reia.2025.202750
- [2] Farzana, S., Sarker, M. S., Chau, T., & Mamun, K. A. (2021). Technological evolvement in AAC modalities to assist verbally challenged ASD children to communicate: A systematic review. *Disability and Rehabilitation: Assistive Technology*, 1–15. <https://doi.org/10.1080/17483107.2021.2005571>
- [3] Cole, R., van Vuuren, S., Pellom, B., Hacioglu, K., Ma, J., Movellan, J., Schwartz, S., Wade-Stein, D., Ward, W., & Yan, J. (2003). Perceptive animated interfaces: First steps toward a new paradigm for human-computer interaction. *Proceedings of the IEEE*, 91(9), 1391–1405. <https://doi.org/10.1109/JPROC.2003.817143>
- [4] Kumar, P. V., Balamurugan, D., Suriya, K., & Reshma, R. (2023). An improved learning experience for English language using big data analytics and AAC. In *Proceedings of the 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*(pp. 347–352). IEEE. <https://doi.org/10.1109/ICSCDS56580.2023.10104864>
- [5] Vogt, P., de Haas, R., de Jong, C., Baxter, P., & Krahmer, E. (2020). Adapt, explain, engage: A study on how social robots can scaffold second-language learning of children. *ACM Transactions on Human-Robot Interaction*, 9(1), 1–27. <https://doi.org/10.1145/3366422>
- [6] Urrea, A. L., Fernández-Torres, V., Rodríguez-Ortiz, I. R., & Saldaña, D. (2024). The use of technology-assisted intervention in vocabulary learning for children with autism spectrum disorder: A systematic review. *Frontiers in Psychology*, 15, 1370965. <https://doi.org/10.3389/fpsyg.2024.1370965>
- [7] Aburukba, R., Aloul, F., Mahmoud, A., Kamili, K., & Ajmal, S. (2017). AutiAid: A learning mobile application for autistic children. In *Proceedings of the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1–6). IEEE. <https://doi.org/10.1109/HealthCom.2017.8210770>
- [8] Lai, S., Deng, C., Zhou, C., Bao, M., Yan, J., Li, H., Yao, L., & Wang, Y. (2025). ASD-iLLM: An intervention large language model for autistic children. In *Findings of the Association for Computational Linguistics: EMNLP 2025*. Association for Computational Linguistics.
- [9] Mishra, R., Welch, K. C., & Popa, D. O. (2024). Human-mediated large language models for robotic intervention in children with autism spectrum disorders. *arXiv preprint arXiv:2402.00260*. <https://arxiv.org/abs/2402.00260>

- [10] Bosker, H. R. (2021). Using fuzzy string matching for automated assessment of listener transcripts in speech intelligibility studies. *Behavior Research Methods*, 53, 1945-1953. <https://doi.org/10.3758/s13428-021-01542-4>
- [11] Pham, T., & Nguyen, H. (2023). Improving the accuracy of speech recognition models for non-native speakers: A case study of Vietnamese elementary students learning English. *Scientific Review*, 9(2), 10-14.
- [12] H. U. Hashim, M. M. Yunus, and H. Norman, "English as secondary language learning and autism spectrum disorder: The obstacles in teaching and learning the language," *Arab World English Journal*, vol. 12, no. 2, pp. 22-30, 2021.
- [13] A. M. Mohamed and T. S. Shaaban, "Promoting multilingual skills in children with autism: Parent and teacher perspectives on teaching ESL," *ResearchGate*, 2024.
- [14] Global Burden of Disease Study 2021 Autism Spectrum Collaborators, "The global epidemiology and health burden of the autism spectrum: Findings from the Global Burden of Disease Study 2021," *The Lancet Psychiatry*, vol. 12, pp. 111-121, Dec. 2024, doi: 10.1016/S2215-0366(24)00363-8.
- [15] K. A. Shaw, S. Williams, M. E. Patrick et al., "Prevalence and early identification of autism spectrum disorder among children aged 4 and 8 years - Autism and Developmental Disabilities Monitoring Network, 16 Sites, United States, 2022," *MMWR Surveillance Summaries*, vol. 74, no. SS-2, pp. 1-22, 2025. [Online]. Available: <https://www.cdc.gov/mmwr/volumes/74/ss/ss7402a1.htm>
- [16] World Health Organization and UNICEF, *Developmental Disabilities Among Children: Neglected Health Needs*. Geneva: WHO, Sep. 2023. [Online]. Available: <https://www.who.int/news/item/15-09-2023-new-reports-highlights-neglected-health-needs-of-children-with-developmental-disabilities>
- [17] C. Siu, "Autism in Sri Lanka," *Annals of Tropical Pediatrics*, vol. 29, no. 4, pp. 327-328, 2009. [Online]. Available: <https://academic.oup.com/tropej/article-abstract/55/6/402/1726615>
- [18] "Increase in autism among children: Over 9,000 identified in Sri Lanka," *Newsire.lk*, Apr. 10, 2025. [Online]. Available: <https://www.newsire.lk/2025/04/10/increase-in-autism-among-children-over-9000-identified-in-sri-lanka/>

11. APPENDICES



Appendix 1- Level 1 Dialogue Word Acquisition Data Flow



Appendix 2 – Level 2 Dialogue Construction Data Flow