Ellie's Adventures: A gamified application to enhance Communication, Emotional, Cognitive, and Motor skills in children with autism.

RP 24-25J-114

Project Proposal Report

Umaira M.M | IT21258312

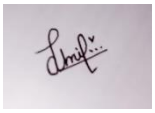
B.Sc. (Hons) in Information Technology, Specializing in Software Engineering

Department of Computer Science and Software Engineering

Sri Lanka Institute of Information Technology
Sri Lanka

Declaration

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Name	Student ID	Signature
Umaira M.M	IT21258312	

Signature of the Supervisor

(Dr. Kalpani Manathunga)

…….

Date

..……

Abstract

This proposal aims to enhance the speech and language development of children with autism by leveraging advanced speech recognition technologies and interactive educational tools. The core component of the proposed system involves a customized, interactive flashcard game that adapts to the individual needs of each child. The system is reinforced by generative AI, which dynamically suggests flashcards based on real-time performance analysis. By integrating datasets of lip patterns and voice recognition, the platform is tailored to improve both speech and language skills. Additionally, the system incorporates a sensory-friendly design and is developed using cutting-edge software solutions, including Google Confidential Computing, multi-tenant encrypted storage, and a layered client-server architecture. This research proposes a novel approach to language acquisition for children with autism, offering a personalized, secure, and effective educational experience.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

Abbreviation	Description
ASD	Autism Spectrum Disorder
WHO	World Health Organization
DTW	Dynamic Time Warping
MFCC	Mel Frequency Cepstral Coefficient
PECS	Picture Exchange Communications
ABA	Applied Behavioral Analysis
gen	Generative
AI	Artificial Intelligence
LSTM	Long Short-Term Memory
ML	Machine Learning
RGB	Red, Green Blue
CNN	Convoluted Neural Network
ResNet	Residual Neural Network
3DCNN	3-Dimensional Convoluted Neural Network
RNN	Recurrent Neural Network
LLM	Large Language Model
VS	Visual Studio
UI	User Interface
UX	User Experience

1. Introduction

1.1 Background and Literature survey

Autism Spectrum Disorder (ASD) is a spectrum of conditions which is classified as a neurodevelopmental disorder [1]. According to the World Health Organization (WHO) every 1 in 100 children has autism [2]. A study [3] in [4] depicts that the three domain characteristics that are mostly shown in ASD are, “social reciprocity, communication and restricted and repetitive behaviors and interests”. Communication can be broadly classified into speech issues, such as articulation and fluency, and language difficulties, including both understanding (receptive) and expressing (expressive) thoughts and ideas. According to [1], almost all children with autism retain difficulty in speech and language. Whereas in Sri Lanka more than 80% of children with autism are having speech related problems [5].

There are much research done and approaches taken to improve speech and language skills in children with autism. Mobile applications and games are most prominent among them. Mobile games, web-based games and serious games have been developed to improve speech and language skills for ASD following different approaches. Webpages such as WhizKidGame [6]. The aim of this game is working with word-object association through the simple use of games to enhance speech and to avoid distraction. When it comes to mobile applications, there are several games according to the two main operating systems which are Android and IOS. Games such as Voice4u [7] and TapToTalk [8] are developed to enhance speech using visualization and concrete images and output audio signals. These applications are available for both Android and IOS operating systems. The GraceApp [9] which is developed for IOS, follows a picture-based communication method, where a set of boards with pictures are stored and are used by the child to express how they feel through the pictures. The Android application ilearn2Talk [10] , follows a picture-based communication followed by speech therapy to improve speech and language skills of children with autism.

When studies are considered, the research [11], which is a computer game to learn and enhance speech problems for children with autism, this study had developed 2 applications to improve speech and language skills in both English and Spanish. The game shows a set of pictures, and the child is asked to pronounce the word. Here voice recognition is developed using the 2 methods

which are DTW and SPHINX4 respectively for each language and detected voice signals are processed using MFCCs for accurate voice detection. The advance level of the game shows uncolored pictures and lets the child pronounce it, the picture gets colored when the pronunciation is accurate. The game consists of a set of images in a frame and once a frame is completed, the game ends.

Another research [4], which is a mobile based learning of picture exchange communication intervention for caregivers of autistic children, this study follows the PECS, for the improvement of speech and language skills. The study reveals the effectiveness of PECS for speech and language development and how it should be used by parents and caregivers of children with autism. Another mobile based learning approach for teachers and caregivers has been discussed in the research [12]. This research uses Applied Behavioral Analysis (ABA) and PECS as learning mechanisms for children with autism and provides a guide on how these mechanisms can be utilized to teach children with autism to improve communication related problems while being at home without reaching the therapy centers.

A serious game approach developed in general for speech disorders and hearing problems has been discussed in the research [13], the study is specifically not for ASD, the speech improvement section here follows a similar approach to the proposed system. This research has developed a game which is controlled by an avatar and provides words for pronunciation. The avatar in the study is controlled by voice commands with a word detection package. The avatar will be collided with an object under a specific word class and the child is allowed to pronounce the word repeatedly and a score between 1 to 100 is given to evaluate the pronunciation. This study has used the Windows UDP Voice Recognition server to detect the child's voice and assess pronunciation

It is clear from the above analysis that none of these solutions fully address the need for comprehensive evaluation techniques. The existing methods lack the integration of advanced technology to accurately assess and improve communication abilities. Therefore, the proposed system aims to fill this gap by utilizing an avatar for communication and a game with flashcards, which will leverage lip pattern analysis and voice recognition for improved speech and language development.

1.2 Research Gap

There are several research done to improve the conditions of Autism considering different effects of autism. When communication is considered, many applications have been developed to improve speech and language skills. Some common features of those applications consist of voice to text, monitoring speech therapy, using picture exchange communication and following a gamified approach. The Research [11] (Research A), [12](Research B) and [4] (Research C) are such applications which have been developed to improve speech and language skills of children with autism.

Research A is a computer game developed to enhance speech problems and verbal communication for children with autism. The developed game consists of a set of colored pictures and the child is asked to pronounce the word in the picture. A set of images are given within a frame and once the child completes the words within the frame the game ends.

Research B is a mobile application which is developed to improve learning skills of children with autism. This includes speech and language development by allowing the child to pronounce a certain word in an image and identifying the accuracy of the pronunciation using the voice recognition techniques. The PECS method of communication is followed in this research.

Research C too is a mobile application developed for teachers and caregiver of children with autism to improve speech and language skills. This follows ABA and PECS. For speech and language development, a set of words is provided, and the child is allowed to read out the word, pronunciation is analyzed using speech recognition techniques.

When the above 3 research papers are analyzed with the proposed system, it is evident that all 3 papers and the proposed system use voice recognition for image identification. Out of the 3, only research A and the proposed system use voice recognition for image identification and only research A and the proposed system follows a gamified approach.

When the overall papers are considered, none of the research have used lip pattern analysis for speech and voice recognition and analysis and none of the researches have used generative AI to

identify the difficulty areas and to generate and provide new words from that area to improve speech and language development which shows the importance and the novelty of the proposed system.

By following a gamified approach, it is expected to maintain the attention and to make the process of learning speech and language very interesting and engaging to the child while analyzing lip patterns for effective communication enhancement.

The following image shows a table of the summary of the above gap analysis for further understanding.

SYSTEM	Voice recognition for image identification	Gamified approach to improve speech	Lip pattern analysis for evaluation	Usage of generative AI to predict more words
RESEARCH A	✓	✓	✗	✗
RESEARCH B	✓	✗	✗	✗
RESEARCH C	✓	✗	✗	✗
PROPOSED SYSTEM	✓	✓	✓	✓

Figure 1 Research Gap Analysis

1.3 Research Problem

According to studies, in Sri Lanka, 7.4% of children between 18 to 24 months have shows red signs of autism defined by the American Academy of Neurology and Child Neurology Society [14]. Additionally, according to this study, every one in 93 children in Sri Lanka is diagnosed with autism [14].

Most children with autism have difficulties such as,

1. Communication, such as speech and language skills.
2. Difficulty in interacting with other people or with surrounding objects and events.
3. Playing with toys and other objects unnaturally.
4. Difficulty in understanding their own emotions and other people's emotions.

Almost all children with autism retain difficulty in speech and language [1]. According to the studies in [5], about 80% of children with autism in Sri Lanka has difficulties in speech development and 40.9% of children are having language comprehension difficulties. More evidently according to [5], compared to the 90s, now, the mental retardation of children with autism has increased from 42 to 50% which highlights the severity of the issue. When communication is considered, while some children may have limited or no speech, some others might struggle with understanding and using language appropriately [2].

To overcome the challenge of education for children with autism, parents are having to take their kids daily to therapy centers. So, the main problem was, "How to bridge the gap between home and therapy centers and help children perform activities at home, without visiting therapy centers daily?". More precisely,

"How to enhance speech and language skills of children with autism, while being at home without visiting the therapy centers?".

When this problem was analyzed in detail it led to the following sub problems, which are,

I. How can we evaluate a child's current level of speech and language skills?
II. Based on what evaluation criteria will speech and language skills be accessed?
III. What will be the main solution to improve speech and language?
IV. How to identify the areas where a child is having difficulties in speech and language?

v. What measures will be taken to improve the identified difficulty areas?

Considering these challenges, the research aims to develop a comprehensive solution that bridges the gap between home and therapy centers for children with autism, specifically focusing on enhancing speech and language skills. By addressing the outlined sub-problems, this study seeks to create an accessible and effective approach to evaluate and improve speech and language development in children with autism within the home environment. This will not only reduce the burden on parents but also provide a consistent, tailored, and interactive platform that supports the unique needs of each child, ultimately contributing to their overall development and quality of life.

2. Objectives

2.1 Main Objectives

The main objective of the research is to bridge the gap between home and therapy centers and to make the task of performing exercises and activities easy at home under the parental supervision. The main objective of the component described from this report is to improve speech and language skills of children with autism by through a set of games and activities while being at home, without needing to go to therapy sessions daily. Thereby, this objective is to be achieved through the gamified mobile application which consist of a 3D avatar as the main communication agent between the child and the application. Specifically, speech and language development is aimed to achieved by a serious flash card game, where the avatar will be showing a flashcard with a picture of a certain word class. The child's voice and lip patterns are to be captured to analyze pronunciation.

2.2 Specific Objectives

To achieve the above main objective is achieved through the below set of sub objectives. They are,

Identifying and evaluating the child's current level of speech and language.

- In order to identify the child's current level of speech and language a mechanism should be developed and according to the identified level, the game should be able to customize.

Lip pattern Identification of the child while pronouncing a word.

- Lip patterns will be used to evaluate the child's pronunciation. The lip patterns while pronouncing the word will be captured using the device camera and the captured data should be filtered and preprocessed to send for analysis. A machine learning model should be trained to extract the lip patterns and to convert the extracted patterns to text.

Developing an interactive flash card game with flash card for evaluation.

- A flash card game is developed under this sub objective, the game will consist of several word classes which are essential to learn speech and pronunciation. Each level of the game will be on a specific word class and the child is allowed to pronounce the word; A reward will be given for a successful pronunciation using positive reinforcement.

Identifying areas where the child is finding difficult to pronounce.

- After the lip patterns are extracted and it has to be sent to another model to analyze and compare to provide a prediction. The captured voice controverted to text, the original text and the text revived by converting the lip patterns will be sent to the LSTM module to analyze and provide a prediction. Based on the prediction score, if the words match and a positive score is given, a feedback will be provided and if a negative score is given, that word class will be identified as a word class with a difficulty to pronounce and will be sent to the next objective below.

Suggest words using gen AI for the identified difficulty areas for improvement.

- The identified difficult word class from the previous objective will be taken here. A model will be trained to generate more words from that specific word class using gen AI and new flashcards for those words will be generated to improve the pronunciation of the child.

3. Methodology

The proposed solution for speech and language development of children with autism is a flash card game communicated through the avatar; The avatar is a customizable avatar where the child can pick according to the child's interest. The avatar will be showing flash cards with readable words and pictures from a certain word class and ask the child to pronounce the word. If the child pronounces the word correctly, a new word from the same word class will be given in the same level. The game will have several levels where each level will consist of words of a certain word class. If the child is finding it difficult to pronounce a certain word, that word class will be identified by the system and more words/ flashcards of that word class will be given until the child improves his or her pronunciation skills.

3.1 System Architecture

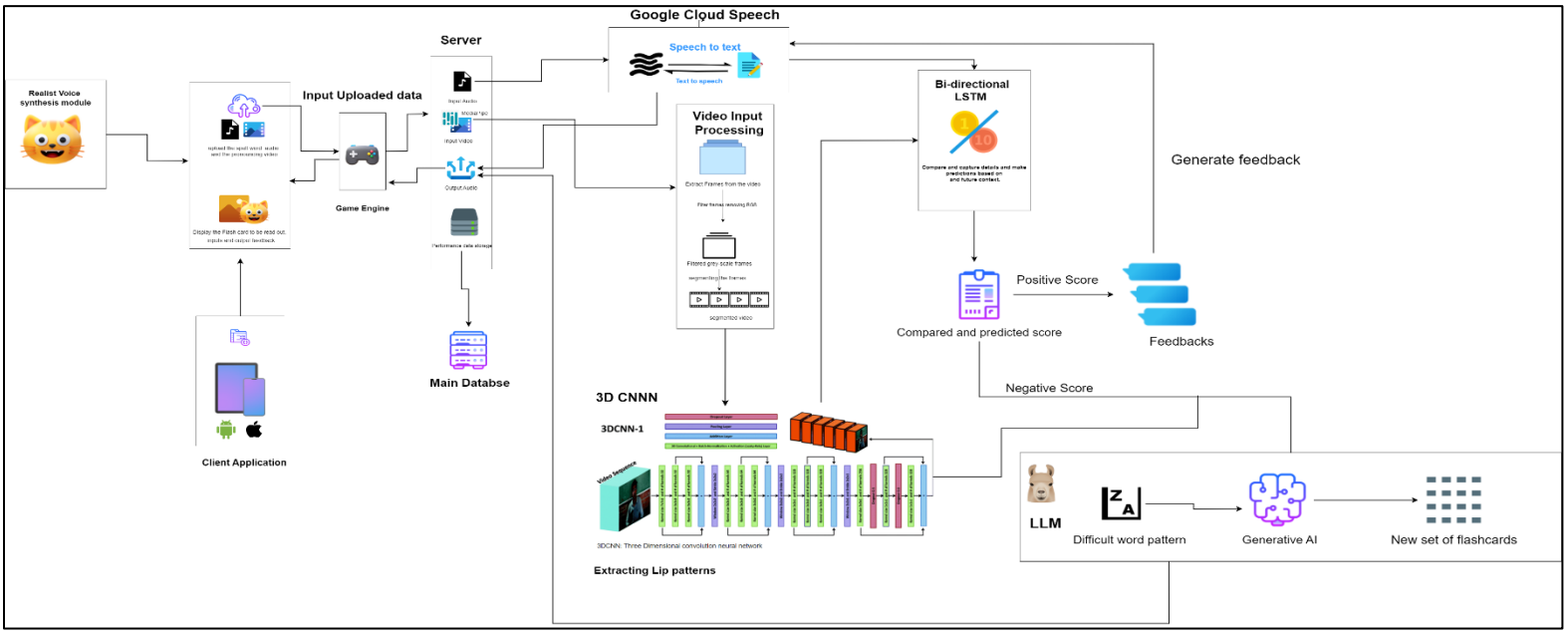


Figure 2 System Overview Diagram

Fig. 2 depicts the high-level architecture of the proposed system. Initially through the mobile application the gaming application has to be opened and the speech and language related flashcard game should be selected. The avatar using the real-time voice synthetic module will be communicating with the child and display flashcards followed by context-based questions.

Next the child will be asked to pronounce the word in the flashcard. While the child pronounces the word, the device camera and the microphone will capture the lip movements while pronouncing the word and the audio of the pronunciation respectively. The captured audio will be uploaded to cloud and through the game engine will be sent to the server. The uploaded audio will be transferred from the server to Google Cloud Speech [15]. This cloud platform is used to transcribe speech to text and text to speech. Therefore, the captured audio pronunciation will initially be converted to text using this platform. The converted text will be temporarily stored for further analysis in later steps. The captured video of pronunciation will be next taken and through the game engine and using the library MediaPipe's face landmark detection [16]. The face landmark detection in media pipe consists of a face mesh which has an Attention Mesh model [17]. Attention in the context of ML is determining the relative importance of each component in a sequence relative to all the others [18]. The attention mesh model relatively applies attention to semantically meaningful face regions, and therefore predicting landmarks more accurately around lips, eyes and irises [17].

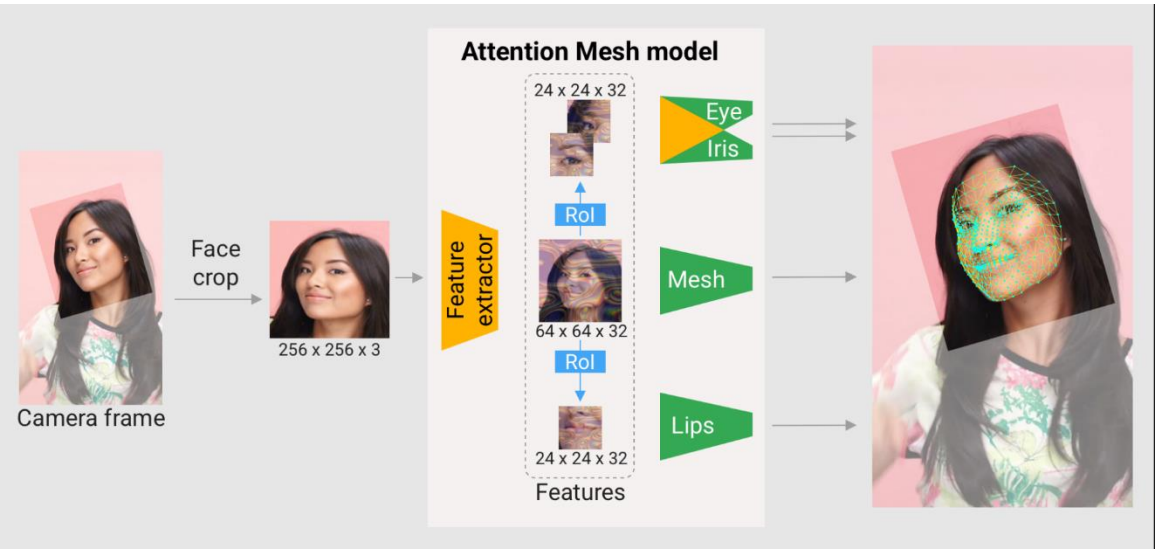


Figure 3 Media Pipe Attention mesh - overview of architecture [17]

The extracted lip features of the video will be processed next. Each video frame will be considered, and its color will be converted from RGB to grey scale using TensorFlow [19]. After the frames are processed, the mean and standard deviation will be computed to normalize the data. Based on the research [20], normalizing and scaling data ensures that the model is not biased towards a certain video frame, which can lead to better generalization and improved accuracy.

The processed sequential video frames will be sent as an input for the CNN model to convert the accurately extracted features to text for comparison. For this research two CNN models which were ResNet and 3D CNN were considered. Compared to ResNet, 3D CNN has the capability to handle both sequential and volumetric data which makes it well-suited for video-based applications like lip-reading. In addition, 3D CNN can directly learn from a dataset without needing to transfer knowledge. According to research conducted on lip reading for hearing impaired children, it is evidently shown that for that in the context of lip reading, 3D CNN is learning at a faster rate than ResNet and has achieved a higher accuracy. Therefore, among the two neural network models, 3D CNN will be chosen for this research. 3D CNN is an extension of 2D CNN which can effectively capture spatial-temporal features of video data. According to [20], 3D CNNs are better suited for the task of lip reading than 2D CNN because it can capture temporal features in addition to spatial features which validates the previous statement. Therefore, using 3D CNN the captured lip patterns will be further extracted, and the pronounced word will be converted to text for further analysis and comparison.

Then the extracted text from lip reading, the text initially converted from the audio and the original text of the word will be sent into an LSTM model for comparison and prediction generation. LSTM is a type of RNN which is used in the field of deep learning. According to the article [21], LSTMs have feedback connections, allowing them to exploit temporal dependencies across sequences of data which makes it more suitable for this task. For this research, a Bi-directional LSTM will be used. Bidirectional algorithm is a commonly used algorithm for lip reading tasks. According to [20], Bidirectional LSTM, involves running the input sequence through two separate RNNs in opposite directions: one is forward and the other in backward. This approach allows the model to consider both the past and future context when making predictions, which is very important for the task of analyzing speech and pronunciation. All three texts (Original text, text obtained from converting the audio, text obtained from converting the lip patterns), extracted from the previous steps will be sent to the bi-directional LSTM and a prediction will be generated according to the comparison of the two inputs with the original text. The texts will be analyzed based on the past and future contexts of the pronunciation and according to [20], the outputs of the two layers will then merge to generate the final prediction.

Based on the prediction given by the LSTM, if the comparison is accurate feedback will be given to the child through the avatar using the real-time synthetic voice model. If the prediction score is low, which means the text comparison is not accurate, the system identifies that word class as a difficulty area for the child. The identified word class is then sent to an LLM model. An LLM is a type of AI which can recognize and generate texts among other tasks [22]. As for this section of the research genAI will be used on top of an LLM. genAI also known as Generative AI, is a type of artificial intelligence technology that can produce various types of content, including text, imagery, audio and synthetic data [23]. For this section of the research, genAI will be used on top an LLM, where the texts of identified difficulty area will be sent to the LLM and more words from that particular word class will be generated, then using genAI, more flashcards for those new words will be generated and sent to the game engine for the further development of the child.

The above-described process will be continued until a child improves their speech and language skills using different word classes in several levels of the game.

Table 1- Technologies, Techniques and Algorithms used

Technologies	Flutter, Python, MediaPipe, TensorFlow, Google Cloud Speech, VS code
Techniques	Text to Speech, Speech to Text, Feature extraction and segmentation, Data augmentation, Data pre-processing
Algorithms	3D CNN, Bi-directional LSTM

3.1.1 Software Solutions

Customization and Personalized UI/UX

Customization and personalized user interface (UI) and user experience (UX) are essential in creating an application that caters to the unique needs of children with autism. By allowing the interface to be tailored to individual preferences, children can interact more comfortably and effectively with the software. Personalized UX helps in creating a more engaging and user-friendly environment, which is crucial for maintaining the child's interest and ensuring that the learning experience is both enjoyable and beneficial. This approach also empowers parents and therapists to adapt the platform to the specific developmental needs and preferences of each child.

Sensory-Friendly Design

A sensory-friendly design focuses on creating an environment that minimizes sensory overload, which is particularly important for children with autism. This includes careful consideration of color schemes, sound levels, and visual elements to ensure they are soothing rather than overwhelming. The design aims to reduce distractions and create a calming atmosphere, making it easier for children to focus on the tasks at hand. By integrating sensory-friendly elements, the application can be more inclusive and accessible, helping children engage with the content without becoming overstimulated.

Google Confidential Computing [24]

Google Confidential Computing is a security technology that ensures data remains encrypted even while being processed in the cloud. This is crucial for handling sensitive information, particularly in applications dealing with children’s data, where privacy and security are paramount. By leveraging this technology, the application can guarantee that personal information, such as speech and language data, remains secure from unauthorized access, thereby building trust with users and adhering to strict data protection regulations.

Multi-Tenant Encrypted Storage

Multi-tenant encrypted storage refers to a storage solution that allows multiple users or tenants to store their data securely within the same system while ensuring that each tenant's data remains

isolated and encrypted. This is particularly important in a system designed for children with autism, as it allows for the secure management of personalized data for multiple users, such as different families or therapy centers. By using encrypted storage, the system can safeguard sensitive information, ensuring that data integrity and confidentiality are maintained across all user accounts.

Layered Client-Server Architecture

A layered client-server architecture is a structural design that separates the application's logic into distinct layers, typically including the presentation layer, application layer, and data layer. This separation allows for better scalability, maintainability, and flexibility of the software. In the context of an application for children with autism, this architecture facilitates smooth communication between the client (user interface) and the server (where data processing occurs). It also enables the system to handle multiple users and large amounts of data efficiently, ensuring a responsive and reliable user experience.

Deploy as a Distributed System

Deploying the application as a distributed system means that it will run across multiple servers or machines, rather than relying on a single server. This approach enhances the system's reliability, scalability, and performance. For a platform designed to support the speech and language development of children with autism, deploying as a distributed system ensures that the application can handle high volumes of data and users, providing a seamless experience even as the user base grows. It also allows for better fault tolerance, reducing the risk of downtime and ensuring continuous availability.

Unit Testing for Flutter Using flutter_test

Unit testing is a crucial part of the development process, ensuring that individual components of the application function correctly. Using `flutter_test`, developers can create automated tests for Flutter applications to verify that each part of the code behaves as expected. This is particularly important in an educational application, where consistent and accurate functionality is essential for the child's learning process. Unit testing helps identify and fix bugs early in the development cycle, leading to a more stable and reliable application.

Linting Code Using Tools Such as Dart Linting for Flutter and PyLint for Python

Linting is the process of analyzing code for potential errors, bugs, or deviations from coding standards. Tools like Dart Lint for Flutter and PyLint for Python help developers maintain high code quality by enforcing best practices and identifying issues early. In a project aimed at enhancing speech and language skills in children with autism, using linting tools ensures that the code is clean, efficient, and less prone to errors. This contributes to the overall robustness and maintainability of the software, ensuring that it remains functional and adaptable over time.

3.2 Commercialization

In developing and commercializing the application, two subscription plans are offered to cater to different user needs: **FREE** and **PREMIUM**. The free plan, available at no cost, provides users with basic features, including limited real-time feedback, allowing families to explore the application's core functionalities without any financial commitment. This plan is ideal for users who are just starting to engage with the platform and want to understand its benefits before making a financial investment.

On the other hand, the premium plan, priced at LKR 4000, offers unlimited access to the full range of the application's features. This includes more personalized feedback, which is crucial for tailoring the learning experience to each child’s unique needs, and expanded access to generative AI features that can further enhance the customization and effectiveness of the platform. The premium plan is designed for families and therapy centers seeking a more in-depth, comprehensive tool to support the development of children with autism. Additionally, by subscribing to the premium plan, users contribute to the ongoing development and improvement of the application, ensuring it remains a cutting-edge resource in autism therapy.

This tiered pricing model not only provides flexibility for users but also supports the sustainability and growth of the application, allowing for continuous innovation and the introduction of new features based on user feedback and technological advancements.

4. Project Requirements and Design

4.1 Functional Requirements

I. The proposed system should be able to recognize speech accurately while the child pronounces the word.
II. The system should accurately capture lip movement while pronouncing.
III. The system should be able to identify the words, the child is finding it to difficult.
IV. The system should be able to predict new words according to the identified difficult areas.
V. The developed games should be short and limited to a certain time span to avoid screen addiction.

4.2 Non-Functional Requirements

I. The system should have user-friendly and sensitive user interfaces since the target user is a child with autism.
II. The speech recognition system should process quickly and provide real-time feedback.
III. Speech capturing should be accurate and reliable to the user.
IV. The system should be secure, since the user is a child, privacy of data should be ensured to the user.
V. The system should be available with high functionality and minimum downtime.

4.3 System Requirements

I. The device used for the application should be a device with a camera and a microphone.
II. The device should have adequate storage and memory to support the system.

III. The device should be compatible with major operating systems.

IV. Should poses relevant dependencies for ML libraries and game development tools.

V. The device should have a string network connection to set up and update the system.

4.4 Personal Requirements

The mobile application will be developed for children with autism who are having difficulties in comprehensively speaking and understand. The games are developed to be performed under parental supervision and with acceptance of caregivers. Data and resources needed for the research are to be provided by the main external supervisor according to the requirement.

4.5 Design (System and Software design documents)

4.5.1 Use Case Diagram

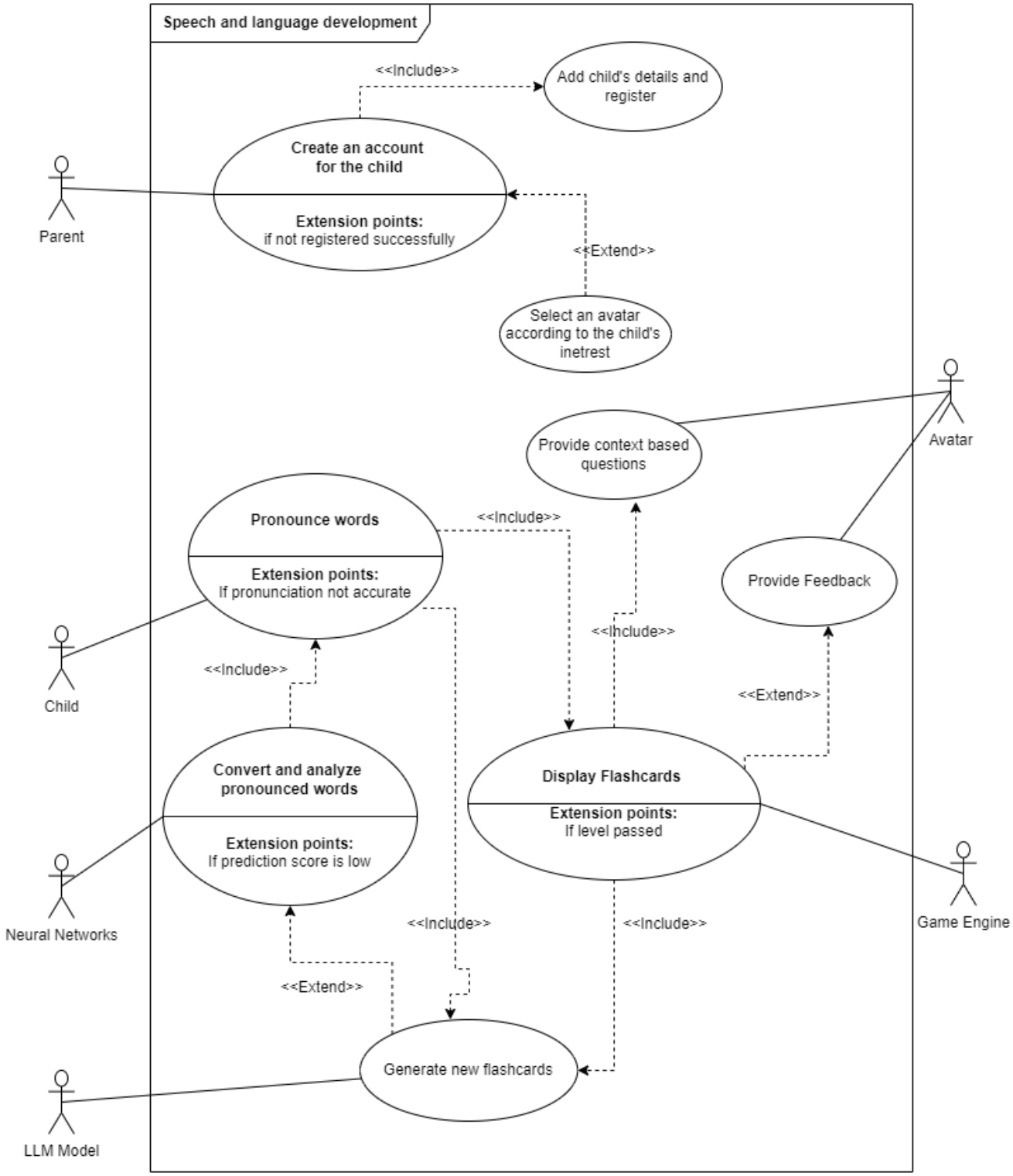


Figure 4 Use case diagram

4.5.2 Tentative UIs

Figure 5 Tentative UIs

5. Gantt Chart

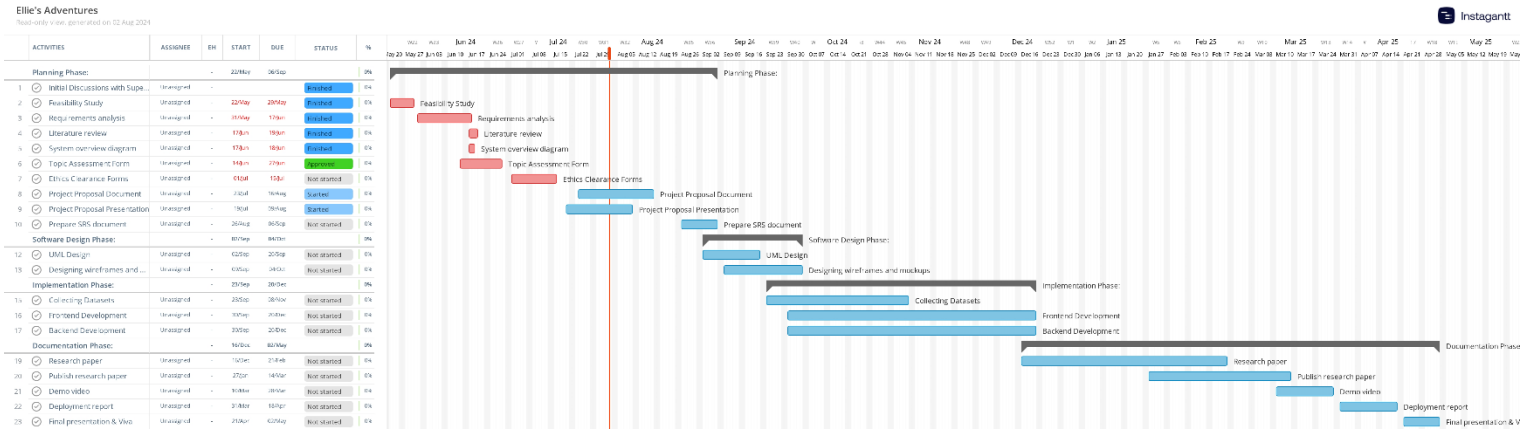


Figure 6 Gantt Chart

6. Work Breakdown Structure (WBS)

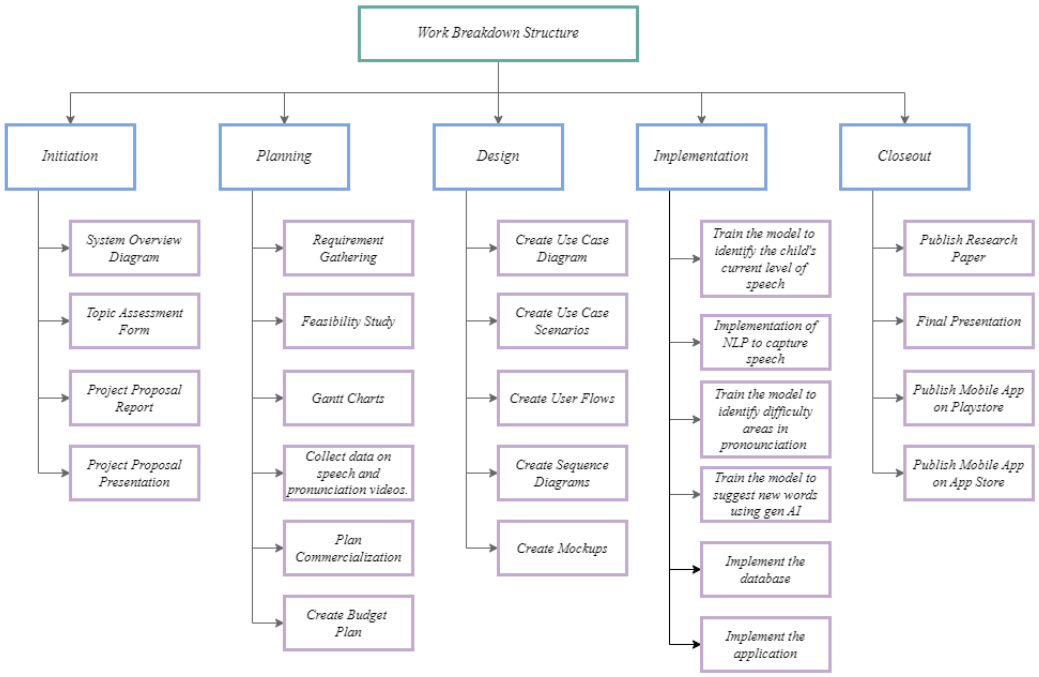


Figure 7 Work Breakdown Structure

References

[1] M. Santiputri, E. B. Sembiring, N. Z. Janah and M. K. Mufida, "Abangmanis: Speech Therapy for Autism Monitoring Mobile Application," in *2019 2nd International Conference on Applied Engineering (ICAE)*, Batam, Indonesia, 2019.

[2] https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders, "World Health Organization," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders.

[3] P.Howling, T.Charman and MM.Ghaziuddin, "The SAGE Handbook of Developmental Disorders," *SAGE,* 2011.

[4] A. S. Shiman, R. A. Adzani, S. Sharif and N. K. Lee, "AutiPECS: Mobile Based Learning of Picture Exchange Communication Intervention for Caregivers of Autistic Children," in *2017 International Conference on Computer and Drone Applications (IConDA)*, Kuching, Malaysia, 2017.

[5] P. H. Perera, "Reasons for which parents seek help for their children with autism," in *Management of Children with Special Needs : Manual for Primary Health Care Workers in Sri Lanka*, Colombo, Sri Lanka., Ministry of Health Sri Lanka, p. 13.

[6] https://www.whizkidgames.com/, "WhisKidGames," [Online]. Available: https://www.whizkidgames.com/.

[7] https://voice4uaac.com/, "Voice4u," [Online]. Available: https://voice4uaac.com/.

[8] https://play.google.com/store/apps/details?id=com.spoken.app&hl=en&pli=1, [Online]. Available: https://play.google.com/store/apps/details?id=com.spoken.app&hl=en&pli=1.

[9] http://www.graceapp.com/what-is-grace-app/, "The Grace App for Autism," [Online]. Available: http://www.graceapp.com/what-is-grace-app/.

[10] www.ilearn2talk.com, "ILearnToTalk," [Online]. Available: www.ilearn2talk.com.

[11] M. Frutos, I. Bustos, B. G. Zapirain and A. M. Zorrilla, " Computer Game to learn and enhance speech problems for Children with Autism," in *2011 16th International Conference on Computer Games (CGAMES)*, Louisville, KY, USA, 2011.

[12] A. S. Shminan, L. J. Choi and S. Sharif, "AutiTEACCH: Mobile-based Learning in a Structured Teaching Approach for Autistic Children Caregivers," in *2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Jakarta, Indonesia, 2021.

[13] N. Nasiri, S. Shirmohammadi and A. Rashed, "A serious game for children with speech disorders and hearing problems," in *2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH)*, Perth, WA, Australia, 2017.

[14] A. H. Vanniarachchy, "Autism in Sri Lanka," Ceylon Today, [Online]. Available: https://ceylontoday.lk/2024/07/20/autism-in-sri-lanka/.

[15] https://cloud.google.com/speech-to-text, "Google Cloud speech-to-text," [Online]. Available: https://cloud.google.com/speech-to-text.

[16] https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker, "Google AI for developers," [Online]. Available: https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker.

[17] https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md, "github.com," [Online]. Available: https://github.com/google-ai-edge/mediapipe/blob/master/docs/solutions/face_mesh.md.

[18] https://en.wikipedia.org/wiki/Attention_(machine_learning), "Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Attention_(machine_learning).

[19] https://www.tensorflow.org/, "TensorFlow," [Online]. Available: https://www.tensorflow.org/.

[20] K. Vayadande, T. Adsare, N. Agrawal, T. Dharmik, A. Patil and S. Zod, "LipReadNet: A Deep Learning Approach to Lip Reading," in *2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC)*, Dharwad, India, 2023.

[21] R. Hamad, "Medium | What is LSTM? Introduction to Long Short-Term Memory," 03 12 2023. [Online]. Available: https://medium.com/@rebeen.jaff/what-is-lstm-introduction-to-long-short-term-memory-66bd3855b9ce#:~:text=Long%20Short%2DTerm%20Memory%20(LSTM,dependencies%20across%20sequences%20of%20data..

[22] https://www.cloudflare.com/learning/ai/what-is-large-language-model/#:~:text=A%20large%20language%20model%20(LLM)%20is%20a%20type%20of%20artificial,network%20called%20a%20transformer%20model., "CloudFlare | What is an LLM?," [Online]. Available: https://www.cloudflare.com/learning/ai/what-is-large-language-model/#:~:text=A%20large%20language%20model%20(LLM)%20is%20a%20type%20of%20artificial,network%20called%20a%20transformer%20model..

[23] https://www.techtarget.com/searchenterpriseai/definition/generative-AI, "TechTarget | What is generative AI?," [Online]. Available: https://www.techtarget.com/searchenterpriseai/definition/generative-AI.

[24] https://cloud.google.com/security/products/confidential-computing, "Google Cloud | Confedential Computing," [Online]. Available: https://cloud.google.com/security/products/confidential-computing.

Appendices

Appendix A – Plagiarism Report

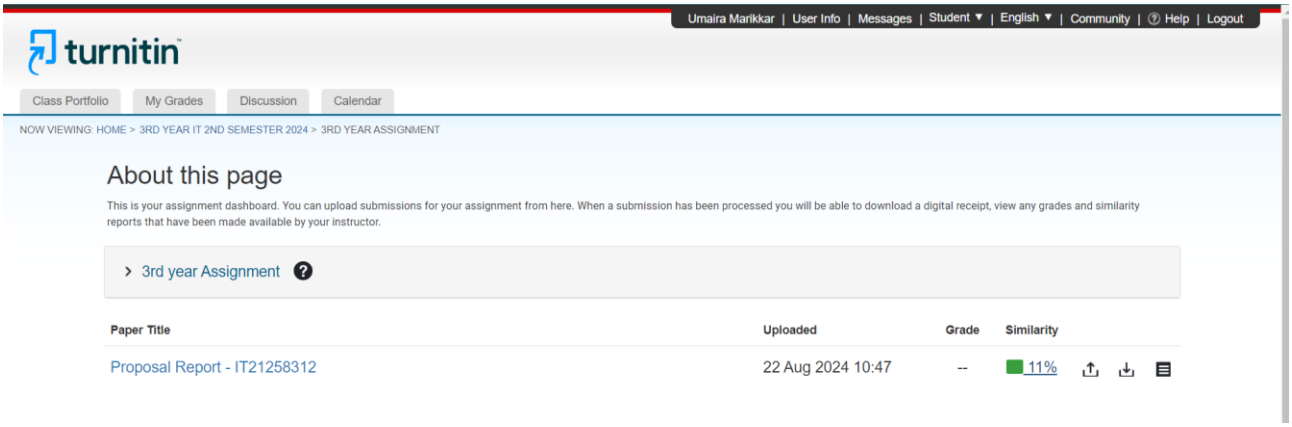


Figure 8 - Plagiarism report