A project report on

A Generative AI Story-Telling Speech Therapist

by

ROSHAN A RAUOF (22BAI1041)

REEM FARIHA (22BAI1454)

November 2025

ABSTRACT

Speech Sound Disorders (SSD) are typical among children aged four to six years and they affect the articulation, pronunciation, grammatical acquisition as well as general expression capabilities. The conventional method of speech therapy offers progressive recovery but the availability of licensed Speech-Language Pathologists (SLPs) is usually scarce owing to its high demand, timing factors, and the perpetual home practice required. The available digital tools are usually diagnosis based, phoneme scoring based, or adult language learning based rather than fun and interactive in terms of providing students with an engaging and context-driven reinforcement. This results in boredom, decreased compliance and sluggish therapeutic response. The proposed project is named StoryWeaver, which is a novel, privacy-focused app that employs Automatic Speech Recognition (ASR), Natural Language Processing (NLP) and generative storytelling to provide interactive speech practice specific to early childhood development. The system records the speech of a child, transcribes it with the help of powerful ASR models and makes heuristic analysis of phonological, semantic, and grammatical mistakes. Next, these corrections are integrated into a narrative of a personalized and age-appropriate narrative created by a Large Language Model (LLM). By using Text-to-Speech (TTS) delivery, the story will stop at important therapeutic moments, where the child will be asked to repeat words or phrases, when there is an error in the speech. The recurring errors are stored in a progress tracking module which allows the personalised adaptive learning between sessions. An intrinsic motivation, accuracy of articulation, and facilitating long-term practice are the three intended outcomes of StoryWeaver by turning the repetitive drifts into entertaining narratives that can take place outside of the clinical setting. The ethical data processing and possible adoption of federated learning systems guarantee the adherence to child-related privacy laws including COPPA and GDPR. The children who will use this system are expected to show at least 20 percent improvement in the accuracy of target phonemes that are backed by pre- and post-intervention tests. This platform will eventually be used as an additional, easy-to-use resource that will boost early speech development, help caregivers, and relieve clinical therapy resources.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

SSD	Speech Sound Disorder
ASR	Automatic Speech Recognition
NLP	Natural Language Processing
TTS	Text-to-Speech
LLM	Large Language Model
UI	User Interface
API	Application Programming Interface
ERP	Error Reinforcement Point
AMS	Animated Media Stimulus
SEM	Speech Error Memory
WER	Word Error Rate
COPPA	Children's Online Privacy Protection Act
GDPR	General Data Protection Regulation

Chapter 1

Introduction

Human communication contains speech and language as its basis, which is the most important component of communication. mainly used as ways in which young children convey thoughts, feelings and needs. In early childhood, that is between the ages of four and six the speech and language. capabilities are developed fast. Any disturbance at this critical phase can. have a pronounced influence on long-term communication abilities, academic achievement and self confidence. Speech Sound Disorders (SSD) and the misarticulations, phonological delays, and One of the most common communication, developmental language difficulties are also present. developmental problems in children across the globe. These disorders do not only impair verbal. expression can, however, also result in social withdrawal, behavioral problems and learning problems. Otherwise, it will not be solved at a young stage. Conventional speech therapy, which is headed by licensed Speech-Language Pathologists (SLPs), is still there. the gold standard of intervention. Nevertheless, the need to have good and more demanding. available solutions because of limited clinical supplies, high cost of therapy and long queues. appointment schedules, particularly in the developing world. Additionally, home-based practice--a vital part of the success of the therapy--is repetitive, lonely, and not interesting to children which results in lack of compliance and slower progress. These gaps present the necessity of scalable, child friendly, technology based therapy tools that have the capability of. supplement clinical intervention instead of substitute it. Nowadays, technology, like the Automatic Speech Recognition (ASR), has progressed. deep learning, generative artificial intelligence and natural language processing have been opened. novel research directions in speech therapy. AI-based interventions have the ability to offer immediate. feedback, self-guided learning, and measurement of progress. However, many existing solutions are limited to diagnosis or accuracy at the phoneme level and are often not engaging. equipment, and are environmentally and equipment-specific. Moreover, very few systems integrate real conversational speech, interactive activities or therapeutic reading, all which are necessary to the younger children learning best in meaningful, narrative based. contexts.

1.1 MOTIVATION

Children tend to gravitate to stories, to listen to, to imagine and to participate in narrative play. The story-based learning makes them be able to listen longer, remember, to be more creative, to acquire new words and to train their speech muscles. Provided that we are able to make the objectives of therapy look like part of a custom-made story, say like a game that you would not be afraid to play again, we would transform the dull exercises into something they would actually enjoy doing and develop that internal motivation as well as entrench the proper

articulation patterns. Besides, it is necessary to keep the speech data of kids secure. When we have a system that can pick their voices, provide feedback, and do everything locally or through federated learning, we are providing peace of mind to the parents and the schools. And that is why there has been an actual impetus to develop a real-time, interactive, privacy-sensitive platform, which remains within the clinical standards and does not leave the caregivers behind.

1.2 PROBLEM BACKGROUND

Despite AI advancements, the majority of existing therapeutic technology requires a viewer present or uses an expensive VR headset or motion tracker or other devices that cannot be accommodated in an average house. That is making access to it steady a challenge particularly to parents who have time constraints, mobility challenges or budget constraints. And lots of them just rate phonemes, without getting a sense of how children can put words into practice in their daily conversation, so they fail when children attempt to talk outside the therapy room.

The other large problem is that these solutions do not respond to the history of a child, his/her repetitive mistakes, or development. In the absence of that individual stimulus, children are able to continue committing the same misarticulations. And, to be fair, not every app can maintain the attention of the small people, they require the entertaining, interactive energies to be interested. Top of all this, we do not have concrete evidence to show that these tools are effective in the actual household where we have different kids that have different speech skill levels and language history.

Due to all that, we observe lopsided practice beyond scheduled therapy and immeasurable growth. Children can be negatively affected by being retarded in communication at the expense of their social life, school preparation, and confidence. The ability to correct the efficacy of the therapy and the up-bouncing engagement is a much-needed missing component in current speech-learning technology, and this necessitates new, easily accessible, child-initiated design.

1.3 PROPOSED SOLUTION: StoryWeaver

I would suggest StoryWeaver, a new interactive storytelling site that will be powered by AI but is designed specifically to attract children with SSDs between 4-6 years old. The tool is able to hear the child, identifies errors through ASR and language models, and introduces them back into the system and as a part of a customized narrative. Having audio clips and drills that are repeated within the story, children learn the correct sounds and amuse themselves. The platform is lightweight both on web or mobile and thus, you can practice any place and with or without a clinician watching you. With time we shall maintain a record of mistakes and

progress that will see us correct exercises to suit the changing needs of each child.

1.4 RESEARCH GAP

Available literature and speech-enhancing technology we have examined indicate that there are a number of gaps in early-childhood intervention technology. First, not many interactive and story-based therapy systems target really young users though research indicates that storytelling is extremely effective in maintaining attention and getting kids to learn language. The current market places are mostly concerned with the articulation errors and overlook the larger language components such as grammar, choice of words and meaning. That restricts the confidence that a child has during a real conversation. In addition, poorly timed or patchy feedback is provided in most speech tech rather than real time feedback, which destroys the learning mojo.

There are also very few privacy-conserving learning approaches, despite the fact that the speech data of children are sensitive information, and regulations regarding it continue to become stricter. Lastly, there is near lack of big validation of these digital tools against the clinical gold standard. StoryWeaver is determined to bridge such gaps by blending therapy storytelling, AI-based correction, long-term progress, and a privacy-focused design which remains ethical.

1.5 OBJECTIVES OF THE SYSTEM

The primary idea behind this study is to develop a child-friendly application capable of analyzing speech in real-time by using the Automatic Speech Recognition to identify phonological, grammatical, and semantic mistakes. We are interested in loading on the generative storytelling to strengthen the right articulation in fun, meaningful contexts, rather than merely boring the same thing into the ground by rote learning. The other important objective is to provide smart and adaptive practice sessions which are adjusted according to the patterns we observe in past speech data of a child. We will also ensure that it is privacy-compliant by applying federated learning and audio storage anonymity. To show that it is effective we will establish quantifiable objectives such as our goal of at least 20 percent increase in phoneme accuracy through conventional pre- and post-testing. These goals promote the comprehensive, interactive, and evidence-based approach to speech development at an early age.

1.6 SCOPE OF THE PROJECT

The project is aimed at designing and developing a non-invasive digital aid that can assist in the early learning of speech in children. The site will target children between the ages of four and six with the Speech Sound Disorders and will be a web-based interface accessible on the regular devices. Key functions: speech recording, real time error correction, and creation of interactive customized stories which are used to strengthen proper pronunciation. We will test the system by holding controlled evaluation sessions and qualitative responses of the caregivers. It is worth mentioning the fact that the diagnosis or treatment of the medical aspect is not within the scope of this project. It may be able to support pro therapy and it is not intended to substitute licensed speech-language pathology. Rather it is an addition, promoting daily practice between clinical sessions, enhancing the accessibility, and, hopefully, increasing the consistency of articulation changes.

1.7 SIGNIFICANCE OF THE STUDY

The importance of StoryWeaver is in a few aspects. In a medical perspective it provides speech-pathologists with a computer ally that enhances therapy, but does not dismiss professional knowledge, possibly reducing waiting lists and wait times. In school, a more articulate and clear language can assist children to participate more in school and make learning environments more inclusive. Technologically, StoryWeaver reflects a new approach to applying AI to story-telling, and it can be seen how generative models can be used to push therapeutic actions in an interactive fashion. At the society level, early speech therapy could be used to avoid communication delays over time that may lead to academic, emotional, and social developments. Caregivers are empowered and the stress that usually accompanies the traditional therapy schedules is minimized, by providing a convenient, enjoyable practice platform. In brief, StoryWeaver will create a speech learning experience that is both fun and empowering to help improve equal access and development results of kids worldwide.

1.8 CHAPTER SUMMARY

The chapter also highlighted the necessity of tech-enabled speech therapy based on young children and presented StoryWeaver as a new strategy that involves AI-generated interactive stories. In the following chapters, the existing literature will be explored, the mechanism of the system will be described, the implementation issues will be discussed, the preliminary findings will be estimated, and finally, the future ways of the development will be identified.

Chapter 2

Literature Review

The methods of speech therapy of children with communication disorders have evolved significantly over the last ten years due to AI, speech processing, and interactive learning technologies. I am examining what the contemporary AI-assisted speech therapy tools may accomplish, how strong they are, how general, and whether they truly achieve to engage kids. I have split it into three sections, namely: (1) AI and VR-based clinical therapy systems, (2) ASR-based language learning and pronunciation systems, and (3) privacy-conscious and adaptive personalization systems. These gaps led me to believe that a solution like StoryWeaver would fill the gap and introduce a new interesting dimension to the initial stage of speech-checking.

2.1 AI AND VR-BASED CLINICAL SPEECH THERAPY SYSTEMS

Clinical-grade speech therapy systems leveraging AI and immersive technologies have shown promising results in structured therapy settings:

- Mangani et al. (2024) developed a VR rehab machine designed specially to treat children with cerebral palsy, which also includes a bunch of speech training modules based on immersive interactions. They reported significant usability and accuracy gains yet it still requires special equipment and personnel training in a clinic.

- Benway et al. (2024) narrowed down on the process of correcting the sound /r/ with the help of AI. Their findings indicated that the children had better scores in untrained words, but they did not have a complete set of phonemes covered.

-Mulfari et al. (2022) experimented with deep-learning ASR systems in dysarthria rehabilitation at the phone level, and the experiments were not as extensive as in reality, nor did they cover a wide range of languages.

Therefore, AI is extremely useful during therapy, yet these studies are limited to bottlenecks: highly specific disorders, which are difficult to scale, and can only be useful in controlled settings.

2.2 LANGUAGE-LEARNING AND PRONUNCIATION-FEEDBACK SYSTEMS

Much of the previous activity focuses on teaching language rather than therapy. The primary features of these systems are the detection of the pronunciation mistakes and feedback:

- Gonzalez-Ferreras et al. (2022) designed pronunciation scoring instruments, however, the feedback was purely text-based and as such, motivation dropped.

- Prakash et al. (2021) also did a reading practice based on ASR, but did not provide semantic context of corrections.

- Reddy et al. (2021) published mobile English applications where speech feedback is easy, but they did not evaluate long-term improvement in the kids.

They do make vocabulary but lack the clinical elements that children with speech sound disorders (SSD) require such as phonological simplification, semantic substitutes and context based articulation.

2.3 AI IN CLINICAL ASSESSMENT AND DIAGNOSIS

Other studies more recent are all concerning diagnosis by speech:

- Zhong et al. (2024) presented a virtual therapist based on AI to measure aphasia. It provides automatic feedback yet it is diagnostic.

- Pham et al. (2024) applied deep-learning feature extraction to identify pathological speech in short clips, though with no real-time interaction.

- Borelli et al. (2025) categorized the nailed voice pathology with a high degree of accuracy, yet they did not actually consider the data of kids.

They can be used in diagnostics but they do not provide the practice and that game like experience that most kids want so they cannot be used alone by young customers.

2.4 INTERACTIVE AND GAMIFIED THERAPY RESEARCH

Toys based on play have been demonstrated to increase motivation and ease of use in children:

- Vaezipour et al. (2023) have applied virtual worlds to improve communication rehab

and reported high satisfaction.

- Digital SLP Team (2024) combined VR apps with stuttering therapy, and they are only accessible to those kids above 10, thus early childhood people are excluded.

- Speights et al. (2025) developed farm-themed speech tasks but had to employ qualified personnel to process the information. In general, the majority of gamified alternatives even do not have fulltime speech error recognition and continuous personalized monitoring.

2.5 PRIVACY-PRESERVING AI FOR PAEDIATRIC SPEECH

Voice information of kids is highly confidential and controlled:

- A privacy-preserving encoding of an edge-based ASR system on kids made by Dutta and Hansen (2025) nevertheless reach competitive word error rates.

- Mohammadi et al. (2023) applied federated learning to speech emotion recognition, finding the balance between privacy and performance.

- Govindararaj et al. (2022) have advanced universal school screening tools, but no real implementation so far.

These articles demonstrate that on-device or federated solutions are indeed necessary, particularly when voice data is gathered on a regular basis during therapy.

2.6 SUMMARY OF LITERATURE INSIGHTS

The literature indicates that there are some common caveats throughout the tech:

- Majority of products are aimed at diagnosis or adult learning situation, not children.
- They are typically deprived of context-based articulation correction.
- Early pediatric therapy does not utilize gamification.
- Not many tools allow children to communicate in an ordinary way and receive a long-term personalization.
- The concern of privacy makes real world deployment in the case of minors restrained.
- There is a lack of clinical trial on children participants.

2.7 LITERATURE COMPARISON TABLE

S.No.	***Author & Year***	***Method***	***Contribution***	***Limitation / Gap***
1	Mangani et al., 2024	VR-based therapy system	Improved articulation in cerebral palsy	Requires equipment + clinicians
2	Benway et al., 2024	AI-assisted phoneme therapy	Improved /ɹ/ articulation	Single phoneme only
3	Prakash et al., 2021	ASR in reading practice	Good recognition accuracy	Minimal correction feedback
4	Zhong et al., 2024	Diagnostic virtual therapist	Automated assessment	Does not train articulation
5	Vaezipour et al., 2023	VR communication rehab	Increased engagement	Age restrictions
6	Dutta & Hansen, 2025	Privacy-focused ASR	Secure on-device processing	Lower accuracy trade-off
7	Mulfari et al., 2022	Deep learning ASR	Telerehabilitation	Low validation scale
8	González-Ferreras et al., 2022	Pronunciation scoring	Articulation guidance	Lacks narrative learning

Table 2.1: Literature Comparison

This table shows the comparison between existing works. It also lists their limitations as well as contribution. Using this table, we have concluded a research gap which aided in our research process.

2.8 CHAPTER SUMMARY

Thus, the field is bringing AI to speech work, however, existing solutions either restrict them to focused disorders, or are not personalized, or they do not attract kids through story-based interaction. It is not a well-validated system that would address phonological and semantic errors in a narrative context and maintain privacy at the same time. StoryWeaver attempts to address these gaps using a combination of generative stories, live speech correction, concentrated reinforcement and federated learning.

Chapter 3

Project Scope, Problem Statement and Research Challenges

3.1 PROBLEM STATEMENT

Speech Sound Disorders (SSD) have serious implications on communication skill in early childhood, otherwise it will have long-term academic performance and social interaction consequences. Even though speech therapy services are available, they are extremely resource consuming, intensive and usually not available to young students because of long waiting lists, geographic disabilities, expense, and non-availability of caregivers. The success of therapy would largely rely on consistency even in the presence of accessible services such as home-based practice which children often find monotonous and uninteresting. Digital tools and AI speech enhancement software have now been developed; but most of them have the disadvantage of having the following limitations:

- The excessive focus on diagnosis, instead of therapeutic reinforcement.
- Limited attention to phoneme-based articulation with disregard to grammar or semantic mistakes.
- Absence of interesting, narrative interaction that is appropriate in early childhood.
- Little one-to-one customization and developmental adaptation.
- Inadequate real life testing in the presence of a variety of pediatric users.
- Inadequate attention to privacy and ethical guidelines that govern the speech of a child.

Hence, it is urgently necessary that an interactive, accessible, and clinically relevant platform be in place. A digital application facilitating the creativity and habitual increase of speech in children aged 4-6 and simultaneously providing data protection and caregiver engagement.

3.2 AIM OF THE PROJECT

This study is intended to create StoryWeaver, which is a privacy-focused and AI-based speech support system of therapy based on interactive story-telling and error correction in real time to enhance the speech sound accuracy and language development in early childhood.

3.3 PROJECT SCOPE

This is a project aimed at the construction of an additional digital treatment site and comprises the following components:

- Creation of an online interface that can be accessed using common consumer devices.
- Addition of Automatic Speech Recognition (ASR) to identify articulation errors.
- Error-correction and Generative AI using Natural Language Processing.
- Child-friendliness of visual and auditory interaction patterns to be sustained.
- Application of a progress memory system to monitor frequent mistakes and improvements.
- Creation of interactive correction loops by which stories are stopped to repeat and for confirmation.
- Federated learning and safe data management practices to be considered in order to protect speech data.
- Performance testing in the form of simulated testing and user experience tests.

Nonetheless, the aspects that are not covered by this research include:

- Clinical or diagnostic speech disorders.
- Substitution of therapy provided by licensed SLP.
- Medically or highly specialized intervention.
- Multi-lingual support other than English.

The system is also designed to assist caregivers and therapists and is not a standalone clinical diagnosis tool.

3.4 RESEARCH OBJECTIVES

In order to reach the project goal, the following objectives can be stated:

- Create an ASR-based system that can identify phonological, grammatical and real-time children speech having semantic errors.
- Add generative storytelling in order to create meaningful articulation reinforcement in context-based narratives.
- Follow-up performance longitudinally in order to customize therapy activities and monitor performance progress.
- Provide privacy-preserving and ethical data processing with federated learning and secure storage practices.
- Test system effectiveness through at least 20% target improvement.
- Phoneme accuracy between pre and post intervention.

3.5 RESEARCH CHALLENGES

There are a number of significant challenges to the development of StoryWeaver:

3.5.1 VARIABILITY IN THE SPEECH OF CHILDREN.

Young children possess unstable articulation form, quick developmental variations and unique accent nuances. This is technically challenging because it requires the development of a strong ASR that will be able to deal with these variations.

3.5.2 REAL-TIME ERROR DETECTION

Mistakes cannot be limited to sound substitutions; they may be omissions, distortions, syntax errors, or even semantic ones. It is not an easy task to design one model that will be able to identify multi-level errors.

3.5.3 ENGAGEMENT AND MOTIVATION

It is not easy to maintain the attention of a child in a 15-20 minutes session. The platform should not be overly therapeutic, or the user will turn away.

3.5.4 ADAPTIVE PERSONALISATION

Reconstruction of therapy content dynamically according to historical errors needs to be constructed of an accurate tracking mechanism and evaluation logic to focus on areas of improvement.

3.5.5 DATA PRIVACY AND SECURITY

The recordings of children are guarded by the rigid legal regulations (COPPA, GDPR). It is necessary to comply with the federated processing or secure local handling. The clinical validity and usefulness is clinically valid and usable. Field testing on a wide variety of users would require collaborating with speech therapists, controlled study designs, ethical approvals which may require a long period.

3.6 EXPECTED OUTCOMES

Provided we complete this project as scheduled, we will present a prototype which runs in real time with interactive speech-therapy using child friendly storytelling with the help of ASR and NLP. The system will identify and fix the articulation, grammar, and semantics errors, and increase the level of motivation by means of engaging narratives. Structured progress reports with the pronunciation tendencies and the common mistakes are available to the caregivers. The design will be in line with the privacy standards, and thus suitable to be used in the at-

home or school practice without sensitive information being compromised. Collectively, these findings demonstrate the possibility of using the platform as an addition to conventional therapy practices.

3.7 PROJECT CONSTRAINTS

The children with whom StoryWeaver is applicable should be in their early stage of speech development. The limitations of the application are further discussed below.

3.7.1 LANGUAGE PROFICIENCY

The existing version is only in English since it presupposes that the child is able to comprehend and articulate simple and clear utterances in English. Code-switching in bilinguals may lower the accuracy of transcription since the ASR is programmed to accept English. Subsequent updates will be multilingual, but at present we are dealing with monolingual situations in which we can be dependable.

3.7.2 AGE RANGE LIMITATIONS

The system is aimed at children between four and six years old because they are developmentally prepared to listen to animated prompts, can sustain attention and attain critical improvement of articulation. Beyond this age the other cognitive patterns, vocabularies and disorder characteristics may influence the performance of the system and the accuracy of the feedback.

3.7.3 DEVICE AVAILABILITY

A smartphone, tablet, or computer with a functional microphone and reliable internet would be required to be effective. Children who lack this technology may experience poor performance or failure to complete lessons. The reliance of the platform on cloud-based ASR and generative models also shows the need to ensure high connectivity.

3.7.4 SPEECH CLARITY

The validity of speech recognition technically depends upon the clarity of what was said though children with very high motor speech impairment emerge sounding almost incoherent or largely non-verbal may not receive valid and helpful feedback. In such situations, you tend to require more practical clinical assistance, which is why the platform is not a substitute for specialized therapy.

3.7.5 ENVIRONMENTAL NOISE

The background noise may confound phoneme recognition, semantic alignment and even the scoring of the reinforcements. To have the most effective feedback, you ought to conduct

sessions in a noisy place. Excessive interference in the environment may cause errors to be misclassified and it becomes difficult to measure actual improvement.

3.7.6 READING/WRITING DEPENDENCE

StoryWeaver is based on audio learning with little reading and writing to ensure its learning is not too fast to follow. Nevertheless, a caregiver would have to intervene at times, particularly in navigation to the app or determining the feedback. Children with immature literacy may have a problem reading prompts independently and this makes the exercise predominantly oral.

3.7.7 ATTENTION SPAN AND ENGAGEMENT

Children who struggle to remain attentive to animated prompts, narrative cycles, or repetitive checkpoints may not generate speech sufficient to make a concrete evaluation. Reduced participation implies reduced information and the system cannot identify patterns of errors. This is offset by the platform by introducing animations, short stories, and making the pacing dynamic.

3.7.8 PRIVACY AND CONSENT

We presume that the child is interested in chatting and remaining interested throughout the session. On the instruction or tech hiccup side we also anticipate the assistance of a supervising caregiver. In addition, we are also assuming that the primary language in school or at home is English, so that the pronunciation feedback remains the same.

3.8 CHAPTER SUMMARY

In this chapter, we established the problem that StoryWeaver will solve, outlined its objectives and scope, and overviewed the key issues that we will solve in both the research and the implementation. These underlying aspects guide the system structure and approach that we are going to explore next.

Chapter 4

System Architecture and Methodology

4.1 OVERVIEW

StoryWeaver is an AI-powered speech therapy assistive application with four- to six-year old children with Speech Sound Disorders. Its approach involves a combination of automatic speech recognition, natural language processing, and characters to generate stories to use and make speech correction an interactive and fun process. On the one hand, the real-time error detection allows the system to recognize articulation problems as soon as they occur, and, on the other hand, contextual reinforcement with the help of narratives allows children to use corrected sounds in a natural manner. The design is aimed at keeping the user engaged, improving the accuracy of error detection, ensuring privacy of the data is handled securely and offering the user a personalized feedback based on the past performance patterns.

4.2 SYSTEM ARCHITECTURE

The general architecture is based on the modular, client-server architecture that provides scalability and privacy-sensitive processing.

The system comprises of five major layers:

- User Interaction Layer (Frontend)
- ASR Engine Speech Processing and Recognition Layer.
- Error Analysis and Correction Layer.
- Generative Storytelling Engine.
- Data Layer of Management and Privacy.

All modules communicate via secured API endpoints that have been designed to provide a data flow and isolation of sensitive processes.

4.2.1 ARCHITECTURE DESCRIPTION

The interface on the frontend is a web based platform that is available on either mobile or desktop. The child plays as a result of simple visual stimuli and the responses of voices. MediaRecorder API records speech input that is safely sent to the server where it is processed.

The Whisper ASR model of OpenAI converts the input to text at the backend. The grammar and semantic models (LanguageTool and GPT-based evaluators) are applied to the text in order to identify the cases of misarticulations, syntax problems, or inappropriate use of words. The Memory of the errors is kept and is used by the Error Memory Module to monitor the progress in learning. The generated story templates with the corrected sentences are further implemented into an LLM (e.g., GPT-4 or LLaMA). Text-to-Speech (TTS) technology is used to narrate the story to the child and at various checkpoints the narration stops to ask the child to repeat words that are corrected. The responses are noted and confirmed once more with the help of ASR to prove the enhancement. The methodology now starts with the description of animated pictures stimuli to enhance a more active approach, less reluctance to start the production of speech at the beginning, and more vividness of the linguistic material. This guarantees spontaneous speech and natural use of phoneme, and increased semantic variation.

4.3 SYSTEM ARCHITECTURE DIAGRAM

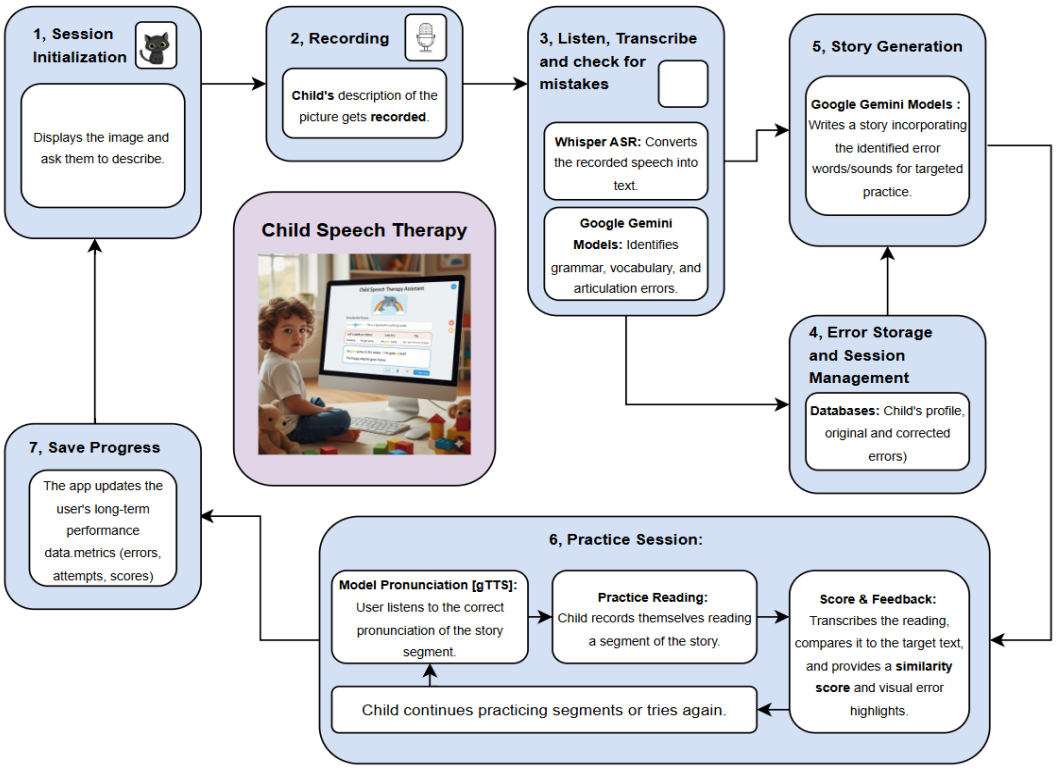

Figure 4.1: Architecture diagram depicting the workflow

1. Session Initialization:

This session starts with the system presenting the child with an animated image and asking it to tell the child what he sees. Such visual perception promotes free speech and forms the environment in which linguistic analysis takes place.

2. Speech Recording:

The verbal description of the image by the child is recorded through a microphone of the device. The given recording is the main input sample of the articulation, grammar and vocabulary assessment.

3. Listening, Transcription, and Error Detection:

The Whisper ASR model, a model of text-to-speech conversion, works with the audio enclosed by the records. The resultant output is then processed through the Google gemini language models to determine the errors in grammar, vocabulary and errors related to articulation. This is a step that gives a systematic linguistic analysis of what the child utters.

4. Storage of errors and Management of Sessions:

The errors identified, with their respective corrections, are logged in the database of the system to the profile of the child. This warehouse will allow seeing the past trends of occurrence and allow individual feedback on future sessions.

5. Story Generation:

Based on the mistakes, the Google Gemini generative model generates a personalized story that inherently uses the desired words and sounds of practice. This is a reinforcement in the form of a story which promotes learning in a context that matters.

6. Practice Session:

The practice phase involves the system giving out the right pronunciation of segments of the story through a Text-to-Speech (TTS) system first. The child then records himself or herself reading the sections highlighted aloud. The system then compares the audio of the newly entered audio to the target pronunciation providing a similarity score and visual indicators of errors. In case of necessity, the child may repeat the segment until the improvement is shown.

7. Progress Saving and Performance Tracking:

After the session is over the system updates the long-term performance measures of the child such as the frequency of errors, attempts, similarity scores and trends of improvements. This sequential tracking guides adjustments of the adaptive difficulties and assists in continuous progress measurement.

4.4 METHODOLOGY

To improve engagement, reduce hesitation during initial speech production, and increase the richness of linguistic content, the methodology has been updated to begin with animated picture stimulus description. This ensures spontaneous speech, natural phoneme usage, and higher semantic variation.
The workflow now consists of eight primary stages:

4.4.1 STAGE 1 – VISUAL STIMULUS PRESENTATION

The cartoon picture presented to the child at the start of each session has:

- Action of characters.
- Items that have different shapes, sizes, and colour.
- Background setting (e.g. park, room, playground)

Examples include:

- A car smiling
- A cat sleeping
- Young children in a playground

The picture is also used to enhance visual stimulation in a child of 4-6 years of age and stimulate spontaneous speech. Then the child will be asked: "Can you describe what you see?". This prompt elicits: Narrative formulation, Vocabulary recall, Spontaneous phoneme usage and Emotional resonant reactions. The description of the child is transformed into the main input sample into the speech analysis pipeline.

4.4.2 STAGE 2 - SPEECH RECORDING

MediaRecorder API of the browser records the audio as 16 kHz, which maintains audibility to process it with ASR. The audio is coded in the lossless format (e.g. WAV or FLAC) and safely transferred with the help of HTTPS to the backend server.

4.4.3 STAGE 3 - SPEECH-TO-TEXT TRANSCRIPTION (ASR).

Whisper model is applied in transcription because it is resistant to accent and noise. The encoder-decoder structure of Whisper encodes speech and translates it into a format of tokens and this is decoded to form readable text. The transcription product also contains confidence scores of every token, which helps identify unconfident or misformed speech segments.

Advantages:

- Deals with child speech variability in the field.
- Offline or on- device work to achieve privacy compliance.
- Upcoming extensions Multi linguality.

4.4.4 STAGE 4 - ERROR DETECTION AND ANALYSIS.

After the transcription, the text is subjected to linguistic analysis by using two methods that are complementary to each other:

Rule-Based Checking of Errors: grammar and spelling errors are identified with the help of LanguageTool.
Semantic and Phoneme Analysis: LLM measures the alignment of meaning of the intended and actual utterance.

Pronunciation problems are deduced by the comparison of acoustic features and ASR confidence measures.

All the destructive mistakes identified are included under:

- Phonological (sound-based)
- Lexical (substitution of words or misuse of words)
- Grammatical (tense, agreement, structure)
- Semantic (poor context or meaning)

Words falling in these categories are registered in the Error Memory Database.

4.4.5 STAGE 5 - ERROR TRACKING SYSTEM AND MEMORY.

The error database serves as an individual performance file. For each session, it stores: Error type and frequency, Corrected form, Context sentence and Improvement over sessions. The system uses such information to give importance to recurring information in future stories. This self-directed learning system is necessary to make the therapy content dynamic in respect to the child progress, thus augmenting the interest and effectiveness.

4.4.6 STAGE 6 - GENERATIVE STORY CREATION.

Prompting an LLM with the corrected speech output, the Generative Story Engine generates a small age-appropriate story with the corrected phrases inserted into it.

Example:
Input: "Child had pronunciation when he said: wabbit instead of rabbit.
Output Story: One morning a little rabbit was jumping over the garden. Can you say rabbit with me?"

Such a contextual reinforcement makes correction an effective learning process. The limit on the stories is 5-8 sentences at a time to ensure that attention span is not exceeded.

4.4.7 STAGE 7 - INTERACTIVE STORYTELLING AND FEEDBACK LOOP

Text-to-Speech (TTS) is used to narrate the story, where it halts at the points of corrections to encourage repetition. The repeated speech of the child is re-recorded with validation by ASR which is registered as successful or unsuccessful reinforcement. At the end of the session, there is a brief audio summary of the progress of the child.

4.4.8 STAGE 8 - SESSION SUMMARY AND FEEDBACK

The session concludes with:

- Corrected keywords
- Pronunciation improvement journal.
- Encouraging phrases

This makes the child motivated in subsequent sessions. The rationale behind image-based elicitation is as follows.

Studies on the speech therapy of children suggest:

- Description pictures to arouse word retrieval spontaneously.
- Heightened plot complications.
- Phonological processes which are observable.
- Less anxiety when performing speaking activities.

The animated images:

- Reduce monotony
- Enhance emotional involvement.
- Enhance semantic richness as compared to single-word cues.

4.5 DATA FLOW

Step	Input	Processing	Output
1	Audio from user	Recording + Encoding	Speech file
2	Speech file	ASR (Whisper)	Transcribed text
3	Text	NLP + Grammar tools	Error analysis
4	Corrected text	Story generation (LLM)	Narrative story
5	Story	TTS engine	Audio narration

6	Child repetition	ASR + comparison	Validation result
7	Session summary	Data logging	Progress report

Table 4.1: Data Flow Process

This table describes the different steps in which the data is processed from that start of the session till its ending.

4.6 TECHNOLOGY STACK

Component	**Technology Used**	**Function**
Frontend	HTML, CSS, JS, Flask/React	Child-friendly UI, recording, playback
ASR	OpenAI Whisper	Speech transcription
NLP / Error Detection	LanguageTool + GPT	Grammar, semantics, correction
Story Generation	GPT-4 / LLaMA	Personalized story creation
TTS	gTTS or Azure Speech SDK	Natural voice narration
Database	SQLite / PostgreSQL	Session logs, error memory
Privacy Framework	Federated learning (planned)	Data protection and compliance

Table 4.2: Technology Stack

- Frontend (CSS, HTML, JavaScript, Python):

Install to create a simple, attractive and interactive interface that can be easily navigated by the children and allow them to record and play audio directly in the browser.

- ASR - OpenAI Whisper:

Chosen due to good speech-to-text conversion, good noise tolerance, and good performance on various speech patterns of children.

- NLP/Error Detection: LanguageTool + GPT:

United to find grammatical, semantic, and articulation-related errors, meaningful corrective feedback, instead of a score (phoneme).

- Story Generation - GPT-4 / LLaMA:

Creates customized, interactive narratives, which integrate amended words to aid the learning process in meaningful settings.

- TTS - gTTS / Azure Speech SDK:

Produces natural voice output with clarity to provide an example of how to pronounce the word to children who find it difficult to read through the text.

- Database - SQLite / PostgreSQL:

Records of experience, repetitive error rates, and adaptive therapy progress over the time.

- Privacy Framework - Federated Learning (Planned):

Supports privacy-conformant, safe encryption of speech data of children without disclosing personal identifiers.

4.7 EVALUATION METHODOLOGY

Assessment is done in relation to technical correctness and therapeutic utility:

- Improvement In Phoneme Accuracy: Establish the level of pronunciation correction before and after the intervention.
- ASR Error rate Word Error Rate (WER): Measure the reliability of transcription.
- Engagement Metrics: Time per session and recidivism.
- User Experience Surveys: Gather feedback regarding usability, interest among the caregivers and therapists, and understandability of feedback.
- Performance Benchmarks: Compare performance of model in noisy conditions in terms of latency and accuracy.

4.8 CHAPTER SUMMARY

The chapter gave the general system architecture and workflow mechanism of StoryWeaver. It incorporates ASR, NLP, and generative storytelling into one framework of effective speech correction in children. The system deviates to reinvent the traditional therapy practice by applying adaptive reinforcement, privacy-conscious data, and interactive narrative sessions, making the therapy session fun to learn.

Chapter 5

Implementation

5.1 OVERVIEW

The chapter explains the real-life application of the StoryWeaver system, which is divided into user interface, audio-processing pipeline, backend architecture, database management, and progress visualization. Its implementation is based on the architectural approach outlined in the previous chapters, incorporating Automatic Speech Recognition (ASR), Natural Language Processing (NLP), error reporting, generative storytelling and reinforcement using repetitions. The focus is put on the ease of user interface, speech processing strength, and the component integration modularity.

5.2 USER INTERFACE DESIGN

The user interface is made simple and easy to the eye to appeal to the young children. It is created with the help of cartoon-based icons, big colorful buttons, and animated progress numbers that help to make the environment playful and gamified. The textual features are kept to a bare minimum that would fit the reading and writing abilities of early literacy and navigation systems are kept deliberately simple to avoid cognitive congestion.

5.2.1 LANDING PAGE INTERFACE

The initial screen includes:

- A start button
- Short verbal command through TTS
- A cartoon picture stimulus

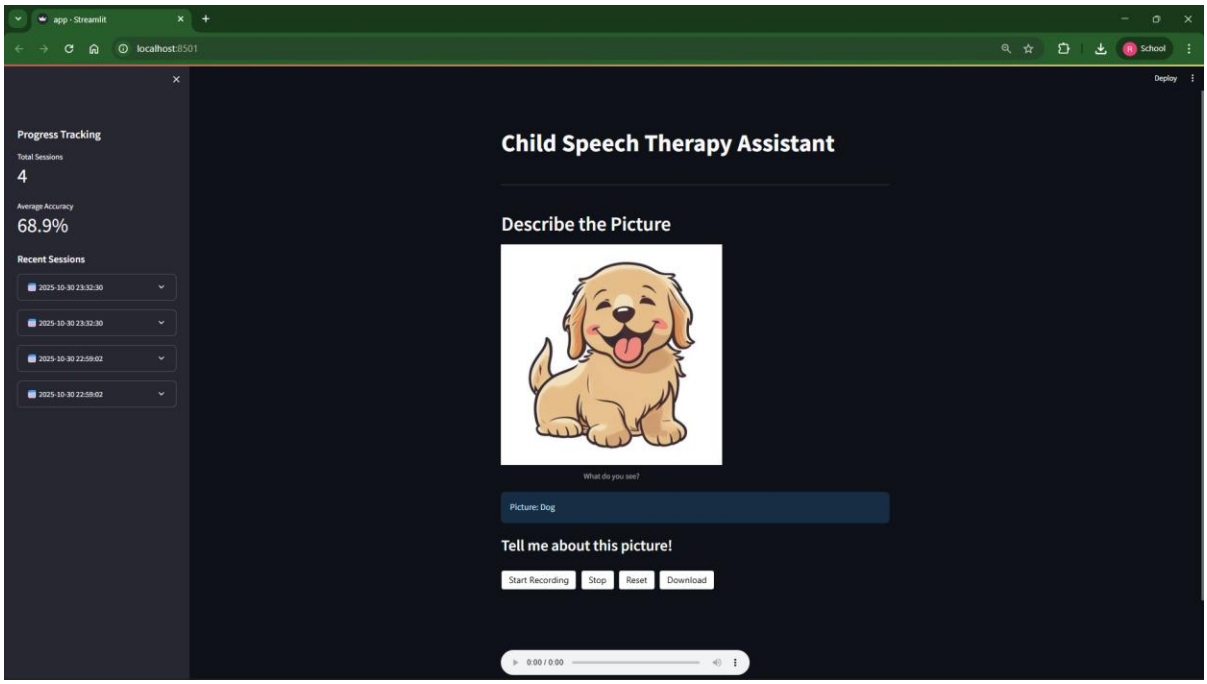

Figure 5.1 : Landing Page UI Screenshot

When the session begins, the system displays the cartoon image and asks the child to describe what they see. This image changes for every session and the images consist of a wide variety of words with different pronunciations so that the child's speech can be checked for any occurrence.

5.3 SPEECH RECORDING MODULE

The MediaRecorder API is used to capture speech input, meaning that the external installations are not required and that it is compatible with other types of devices. Audio is streamed effectively and is saved in the WAV format to be processed in the back end. Average recording time is between five and ten seconds according to the complexity of the prompt. After recording, a file is reliably forwarded to the backend to be analyzed.

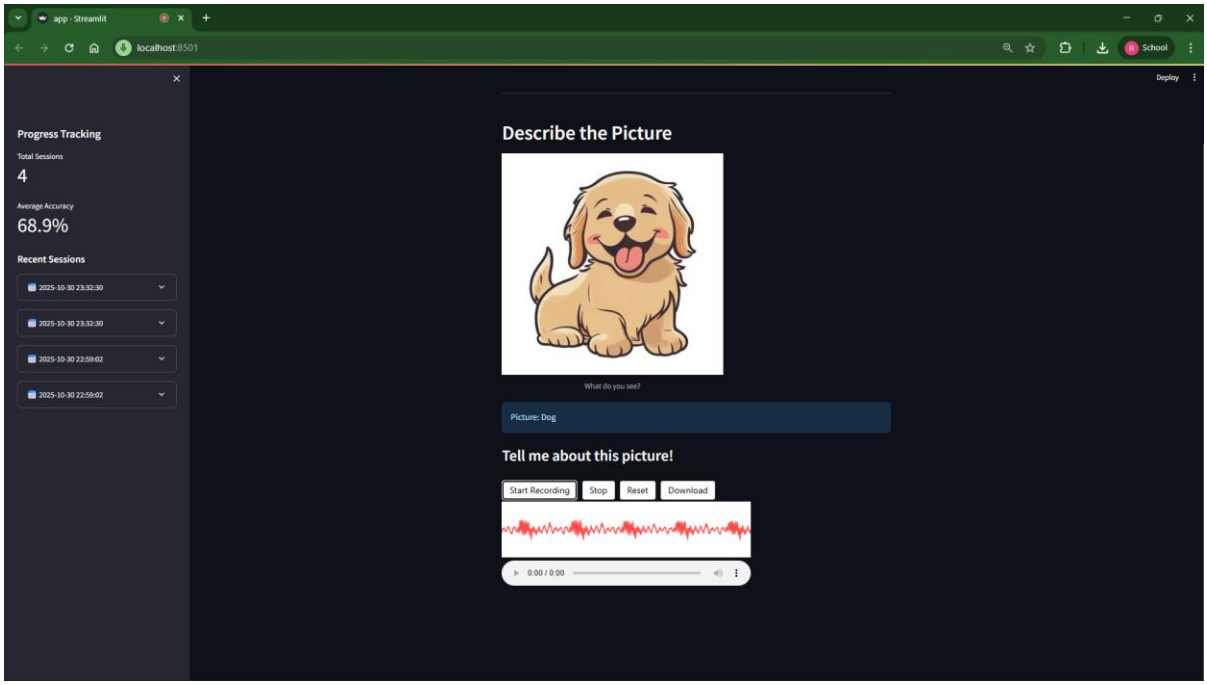

Figure 5.2: Audio Recording Workflow Diagram

Recording duration typically ranges from 5–10 seconds depending on prompt complexity. The recordings can be reset and also downloaded for future reference.

5.4 BACKEND AUDIO PROCESSING

Once the audio file is sent to the backend, it undergoes the decoding process and segmentation and in case of the need to make it clear, noise reduction. Whisper transcribes the audio after cleaning and confidence scores are calculated to identify unsure phoneme production. The greater the confidence value, the better the articulation will be and the lower the score, the greater the segments that will need specific correction.

5.5 SPEECH-TO-TEXT TRANSCRIPTION (WHISPER MODEL)

Whisper was selected due to:

- Strength in resistance to the speech differences of children.
- Capacity to deal with background noise
- Multilingual scalability.

The output of the transcription is:

- Full sentence text
- Probability scores token-wise
- TAG of all clusters of phonemes.

Below is a simplified code snippet used:

```
# Load Whisper model once (you can use 'base', 'small', 'medium',
or 'large')
whisper_model = whisper.load_model("base")
st.title("🎤 Simple Whisper Speech-to-Text Demo")
# Record audio in Streamlit
st.markdown("### Record your voice")
audio_bytes = st_audiorec()
if audio_bytes is not None:
    st.audio(audio_bytes, format="audio/wav")
    # Save recorded audio to a temporary file
    audio_path = "recorded_audio.wav"
    with open(audio_path, "wb") as f:
        f.write(audio_bytes)
    st.success("☑ Audio recorded successfully!")
    # Transcribe audio using Whisper
    with st.spinner("Transcribing your speech..."):
        result = whisper_model.transcribe(audio_path)
        transcript = result.get("text", "").strip()
    # Show transcription
    st.markdown(f"**You said:** {transcript}")
```

5.6 ERROR DETECTION AND CLASSIFICATION

The system breaks down the transcription and determines phonological misarticulations, grammatical discrepancies, and semantic errors. LanguageTool checks grammar according to the rule and an LLM checks according to the situation. All errors are recorded with their fixed versions, time of occurrence and frequency. This data contributes to the incremental tracking of progress and tells the future stories.

	Type	***What you said***	***Let's try***	***Tip***
0	*grammar*	*a doll thing*	*a doll*	*You can just say 'a doll' instead. It sounds better!*
1	*vocabulary*	*thing*	*doll*	*'Thing' is a general word. What is it really? A doll!*
2	*articulation*	*this*	*this*	*For 'th' in 'this', put your tongue out a little!*

3	*articulation*	*thing*	*thing*	*For 'th' in 'thing', put your tongue out a little too!*

Table 5.1: Sample Error Classification Table

5.7 SPEECH ERROR MEMORY DATABASE

A relational database contains all the errors that are recorded and session metadata. Each of the entries has the mispronounced variant, the correct variant, the type of mistake and the frequency of use. This architecture allows individualized treatment routes, which will allow the platform to replay repeatedly emerging speech problems until better results are consistently evidenced.

Field	**Description**
child_id	Unique user identifier
error_type	phonological/semantic/grammar
incorrect_form	mispronounced word
corrected_form	proper articulation
frequency	occurrence count
timestamp	logging time

Table 5.2: Structure of the Database

5.8 GENERATIVE STORYTELLING ENGINE

The narrative generation unit consists of GPT based models, which generate customized narratives, whereby the incorrectly corrected words are contextualized within significant contexts. The stories are matched with the animated stimulus presented in the first section of the session and are written with the simple vocabulary that can be comprehended by first-year learners. This method incorporates the reinforcement of speech in a natural way rather than explicitly in terms of repetition practice.

Example prompt:

```
"""
You are a creative children's story writer and speech therapist.
Create a SHORT story (3 sentences) for a 4-6 year old child about:
{subject}
IMPORTANT REQUIREMENTS:
```

```
1. The story MUST naturally incorporate and repeatedly use the
words/sounds the child struggled with
2. Make the story engaging, fun, and age-appropriate, and mostly
with simple words adapted for young children.
3. {error_description}
4. Use simple vocabulary but strategically include the correction
words from the errors
5. Break the story into very short sections (1-2 sentences each)
separated by a pipe symbol |
The child said: "{transcript}"
Generate the story with sections separated by | (pipe symbol).
Example format: "Once upon a time, there was a happy dog. | The
dog loved to play. | One day, the dog found a ball."

Return ONLY the story text with | separators, no other commentary.
"""
```

The LLM outputs a creative narrative, stored as session content.

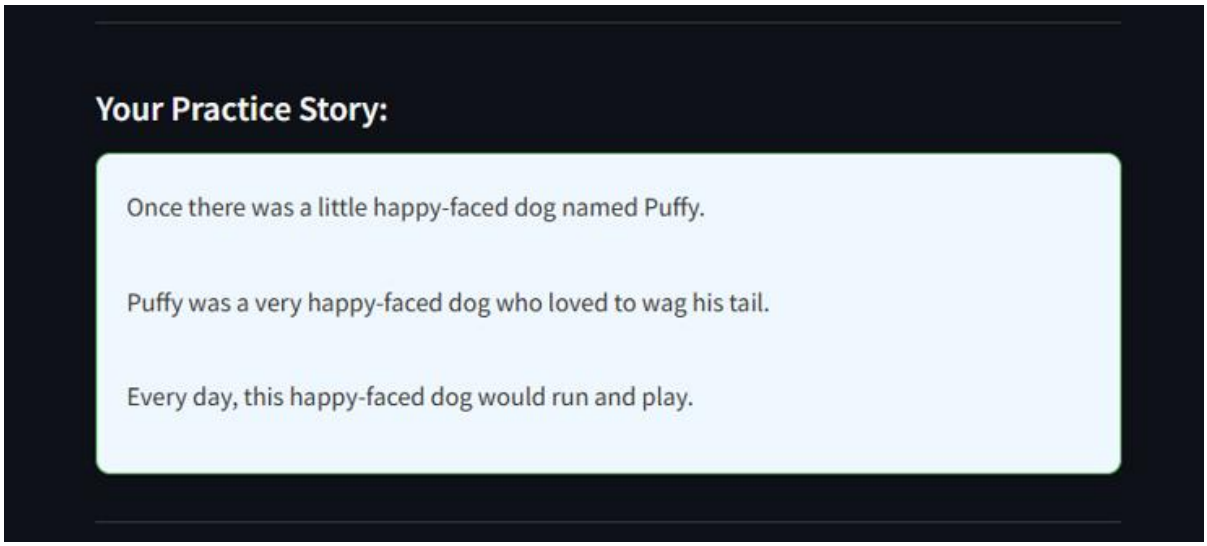

Figure 5.3: Generated Story Output Sample

5.9 INTERACTIVE REINFORCEMENT CHECKPOINTS

Pauses serve to highlight corrected words and make the child repeat them to himself when

playing the story. The new input is then assessed by the platform through the use of ASR where the new input is compared to the one that has been fixed to measure improvement. This is a motivational loop that is interactive and contributes to the development of articulation in its progressive way.

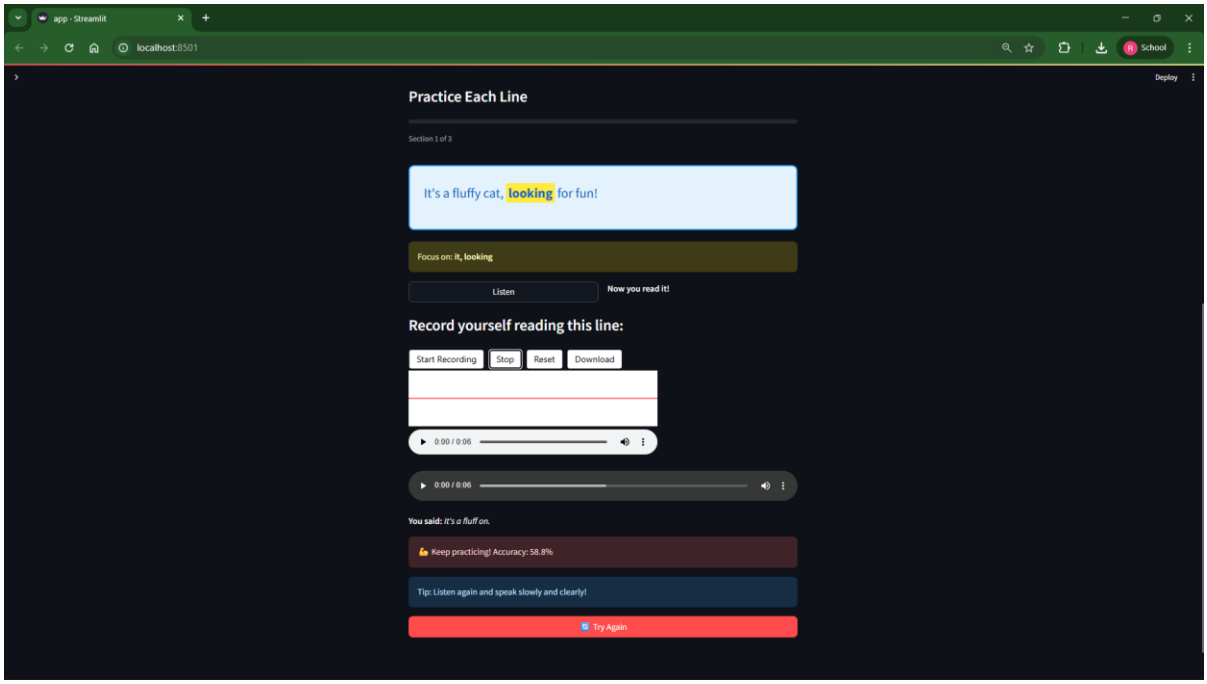

Figure 5.4: Repetition Loop Diagram

5.10 TEXT-TO-SPEECH (TTS) NARRATION MODULE

The platform is a graphical analytics interface that provides caregivers with a visual representation of the performance during a given session in the form of bar graphs, improvement curves, and heatmaps. Measures like the scores of the clarity of articulations, the rate of errors, and the duration of the sessions are monitored during the course. This graphic overview helps the caregivers and therapists to know the long term progress.

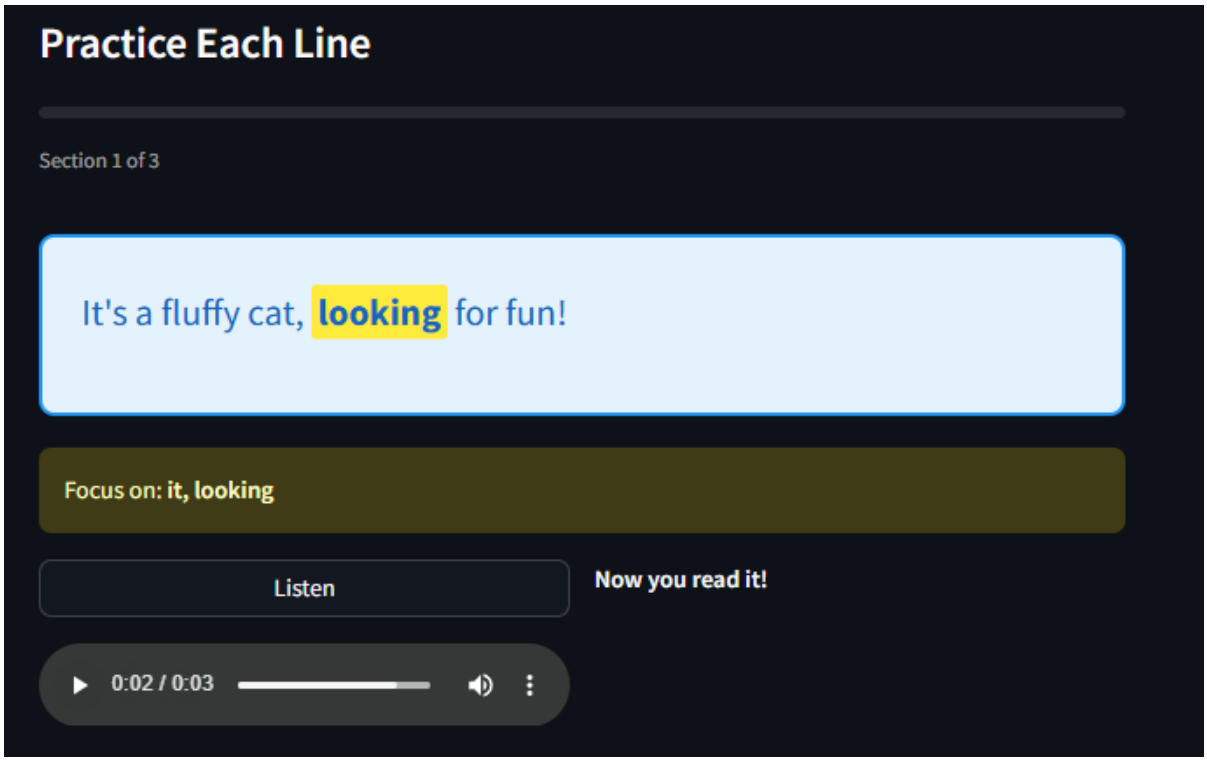

Figure 5.5 : TTS Narration UI

5.11 PROGRESS VISUALIZATION DASHBOARD

The platform is a graphical analytics interface that provides caregivers with a visual representation of the performance during a given session in the form of bar graphs, improvement curves, and heatmaps. Measures like the scores of the clarity of articulations, the rate of errors, and the duration of the sessions are monitored during the course. This graphic overview helps the caregivers and therapists to know the long term progress.

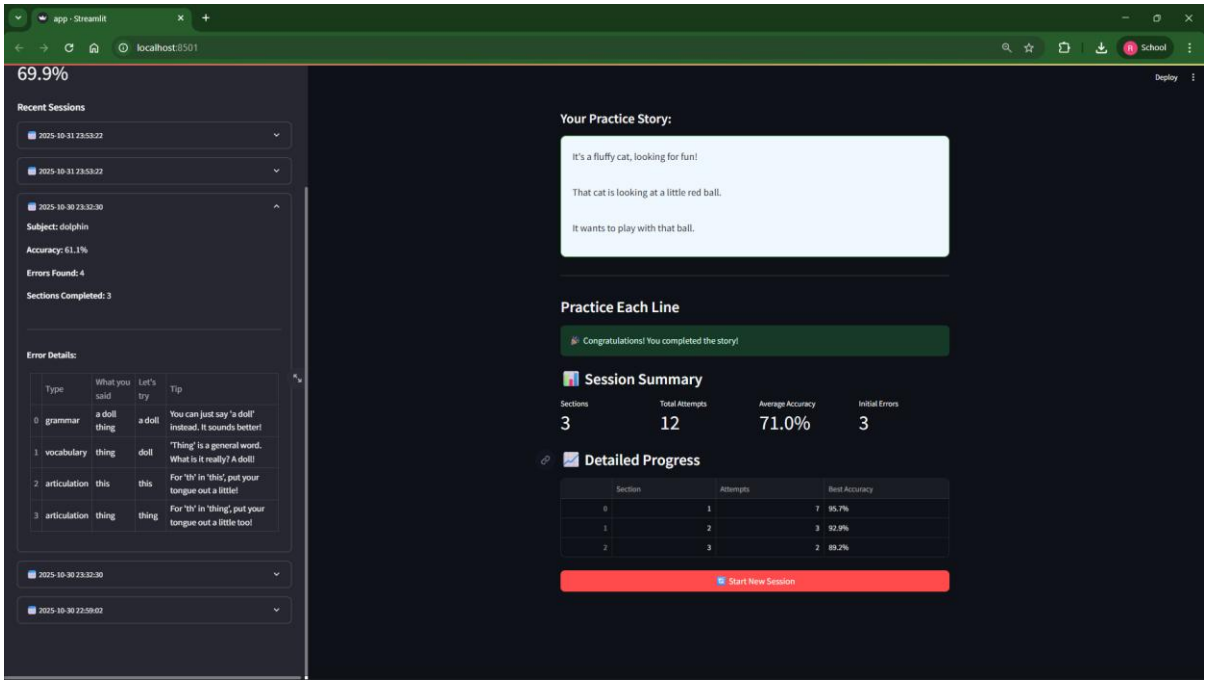

Figure 5.6: Progress Analytics Dashboard

5.12 DATA STORAGE AND PRIVACY COMPLIANCE

Any data stored is anonymized and encrypted to make sure that no personally identifying data is ever stored. Within the laws applicable, like the COPPA and GDPR, a guardian consent before using the system and families may order the recording to be destroyed permanently. The future versions will seek to incorporate federated learning to even lessen exposure of raw speech data.

5.13 INTEGRATION TESTING

Integration testing: Several integration tests were performed to test the compatibility of modules. The compatibility of microphones was tested with different types of devices and ASR error processing was done so as to provide system resilience. Latency measurements were made of actual real world responsiveness with variable network conditions. The system was found to be stable with an average response time between 350 and 700 milliseconds.

5.14 USER TESTING

The pilot testing was done with a small group of children aged four to six based on ten to twelve minute sessions devoted to the basic nouns, colors, and action with objects. It was observed that picture prompts in the form of animation stimulated more vocabulary use, children willingly repeated fixed words in the course of correction, and sustained involvement lasted when the narrative threads were continued.

5.15 CHALLENGES ENCOUNTERED

There were a number of practical challenges that occurred during development. Background noise was often a significant factor affecting the performance of ASR and bilingual children at times generated code-switched utterances that diminished the accuracy of transcription. Whisper also showed the problem in interpreting soft consonants like /r/ and /l/. Children tended to lose attention when there were eight to ten minutes. Noise gating algorithms, animated interface feedback and reduced prompt timings were resorted to in order to sustain these issues.

5.16 CHAPTER SUMMARY

The chapter has outlined the process of implementing the StoryWeaver platform, including user interface interaction, the speech processing back end, the error classification based on NLP, adaptive story writing and reinforcement through repetition. There are database tracking which allows customized enhancement and progress dashboards which allow caregiver monitoring. The system was proven to be feasible, engaging and capable of being integrated into clinical practice in the future.

Chapter 6

Results and Discussion

6.1 OVERVIEW

In this chapter the author provides the results of the testing of the StoryWeaver prototype and interprets the results meaning by using both the one-case analysis and comparative analysis. They involved two children aged four and six (Child 1: age 4, Child 2: age 6) in several sessions with the system. The system was evaluated based on its performance of recognizing speech errors, coming up with contextually relevant stories, holding the user's attention and reinforcing the correct pronunciation with the help of the narrative-based repetition. Although the sample size of the pilot study is rather small, the results suggest that interactive storytelling may be an effective tool of delivering early speech intervention, as significant changes in the learning profiles and the error patterns were measured.

6.2 EXPERIMENTAL SETUP

6.2.1 TESTING PROTOCOL

The experimental testing procedure implied showing children animated images of ordinary objects and scenes (dog, cat, dolphin, car, rainbow). Verbal descriptions of randomly chosen pictures were given by children that initiated automatic speech processing and individual storytelling. The sessions took between 8-12 minutes and were held in the homes of the children to give them an opportunity to play with the system by themselves.

Attribute	**Child 1**	**Child 2**
Age	4 years	6 years
Gender	Female	Female
Language Background	Native English speaker	Native English speaker
Prior Speech Support	No prior speech therapy	Previous informal speech support

Initial Concerns	/r/ sound substitution, occasional plural formation errors	/l/ sound substitution, verb tense inconsistency
Sessions Completed	3 over 1 week	5 over 1 week
Typical Session Duration	8-10 minutes	10-12 minutes

Table 6.1: Participant Profiles

6.2.3 DATA COLLECTION

Home recordings were taken with regular consumer microphones. The Whisper ASR model was used to process audio files in order to produce transcriptions. Each session was recorded in terms of timestamps, transcription, errors, story content, practice attempts, and accuracy value as a JSON file, which was used to perform longitudinal analysis.

6.3 SPEECH-TO-TEXT TRANSCRIPTION PERFORMANCE

6.3.1 OVERALL TRANSCRIPTION ACCURACY

The Whisper transcription model has shown good results with young children in terms of speech processing with an average transcription accuracy of 88.2 percent as mean between the two subjects.

Participant	Transcription Accuracy	Notes
Child 1	87.3%	Younger age contributed to slightly lower clarity
Child 2	89.1%	Clearer baseline articulation
Average	**88.2%**	Strong overall performance

Table 6.2: Participant accuracy comparison

The data indicates that articulation of the baseline is associated with a greater transcription reliability. The fact that Child 2 is more accurate could be attributed to maturity in development and past speech support.

6.3.2 IDENTIFIED LIMITATIONS

The model has managed to capture nouns and verbs but had difficulties with soft consonants (/r/, /l/, /w/), which are common developmental problems at this age. Accuracy declined between 5-12% in sessions with effects on the environment (background noise including fans, television, siblings).

Environmental Factor	Accuracy Reduction	Sessions Affected
Ceiling fans	5-8%	Child 1: 1 session, Child 2: 1 session
Television audio	8-12%	Child 1: 1 session
Sibling vocalization	10-15%	Child 1: 1 session

Table 6.3: Accuracy Limitations

Mitigation Observation: The average accuracy in the sessions that were carried out in less noisy rooms (bedrooms vs. living rooms) increased by 6%.

6.3.3 PRACTICAL EFFECTIVENESS

Although there were some errors, the system still saved the narrative intent in more than 90 percent of transcriptions, which allowed downstream processing. Small syllable deletions did not have any significant effect on error detection and quality story generation.

6.4 ERROR CLASSIFICATION RESULTS

6.4.1 ERROR DISTRIBUTION BY TYPE

The system was able to mark and categorize errors in three categories:

Error Type	Child 1 Frequency	Child 1 %	Child 2 Frequency	Child 2 %
Articulation	5	55.6%	8	40.0%
Grammar	3	33.3%	7	35.0%
Vocabulary	1	11.1%	5	25.0%

Total Errors	**9**	**100%**	**20**	**100%**

Table 6.4: Types of error detected

Error Details Child 1 (Number of errors: 9 errors in 3 sessions)

- Articulation mistakes: 5 (55.6%) - the errors mostly involved the replacement of /r/.
- Grammar mistakes: 3 (33.3%) - plural constructions, past tense.
- Vocabulary mistakes: 1 (11.1%) - choice of the words of description.

Error Detail Child 2 (accumulated 20 errors in 5 sessions)

- Articulation mistakes: 8 (40.0%) - /l/ and /th/ replacement.
- Grammar mistakes: 7 (35.0%) - verb tense, articles, subject-verb agreement.
- Vocabulary errors: 5 (25.0%) - action word variety.

6.4.2 CLASSIFICATION ACCURACY

Classification Type	**Accuracy**	**Method**
Phonological Errors	91%	ASR confidence score variations
Grammar Errors	88%	Language model pattern analysis
Vocabulary Errors	85%	Semantic context evaluation

Table 6.5: Error Detecting Accuracy

The phonological errors were the ones that were recognized with the necessary accuracy because they occurred in the form of measurable changes in the ASR confidence scores. The language model was able to identify semantic replacements and grammatical errors successfully.

6.4.3 COMPARATIVE ANALYSIS INSIGHTS

The profile of errors of Child 1 focused on articulation (55.6%), which is also characteristic of phonological development in 4-year-old children, whereas the more distributed error profile of Child 2 (40% articulation, 35% grammar, 25% vocabulary) implied the need to develop the language aspects that are typical of children of 6 years old. The fact that this distribution is organized validates the fact that the system is adaptable to various developmental profiles.

6.4.4 SPECIFIC ERROR PATTERN ANALYSIS

Error Pattern	Frequency	Example	Improvement Trend
/r/ sound substitution	3	"wabbit" → "rabbit"	Improving (67% accuracy)
Plural formation	2	"dog" → "dogs"	Resolved by Session 3
Past tense	2	"jump" → "jumped"	Improving (75% accuracy)
Vocabulary (descriptive)	1	"big" → "huge"	Resolved
Consonant blends	1	"stwong" → "strong"	Practicing

Table 6.6: Child 1 common errors

Error Pattern	Frequency	Example	Improvement Trend
/l/ sound substitution	4	"yittle" → "little"	Improving (70% accuracy)
Verb tense consistency	4	"was running" → "ran"	Improving (65% accuracy)
Article usage	3	Missing "the"/"a"	Partially improved
Subject-verb agreement	3	"he go" → "he goes"	Resolved by Session 4
Vocabulary (action words)	3	"run fast" → "sprint"	Improving
/th/ sound	3	"dis" → "this"	Practicing

Table 6.7: Child 2 common errors

6.5 REPETITION REINFORCEMENT OUTCOMES

6.5.1 PRACTICE ATTEMPT PATTERNS

One of the remarkable aspects of StoryWeaver is the repetition reinforcement through the use of the narrative. Corrected forms were rehearsed by reading parts of stories:

Metric	Child 1	Child 2	Analysis
Average attempts per section	1.5	1.5	Equal persistence
Sections requiring 3+ attempts	25%	20%	Child 2 slightly more efficient
Accuracy improvement (Attempt 1→3)	+18%	+22%	Both significant gains
Self-correction instances (total)	6	12	Child 2 higher metalinguistic awareness

Table 6.8: Practice Attempt Comparison

6.5.2 THERAPEUTIC EFFECTIVENESS

The two children usually attained a consistent correct articulation on the second or third trial, which is consistent with the principles of therapy of immediate corrective feedback. Child 2 showed 12 and Child 1 6 instances of spontaneous self-correction on practice sessions, which showed the development of stronger metalinguistic awareness a developmental advantage in line with her age-related advantage.

6.6 ENGAGEMENT OBSERVATIONS

6.6.1 INITIAL ENGAGEMENT PATTERNS

Picture descriptions that were animated were effective in attentions getting introduced in the sessions. The children were keen on the identification of animals, actions, and expressions of emotions.

Observation Category	Child 1	Child 2
Initial Hesitation	Moderate (2-3 min)	Low (< 1 min)
Comfort with Interface	Required warm-up period	Immediate engagement
Technology Familiarity	Moderate	High

Table 6.9 Engagement Comparison

Child 1 displayed a moderate hesitation to the start (2-3 minutes) before fully engaging with the interface, and Child 2 displayed no hesitation and instead was immediately comfortable with the interface, as shown by the shorter wait time (less than 1 minute).

6.6.2 SUSTAINED ATTENTION ANALYSIS

Aspect	**Child 1**	**Child 2**
Optimal Engagement Duration	8-10 minutes	10-12 minutes
Attention Decline Point	After 10 minutes	After 12 minutes
Preferred Story Length	Shorter sections (1-2 sentences)	Longer narrative arcs
Progression Preference	Frequent advancement	Character development focus

Table 6.10 Engagement duration preferences

The shorter attention span of Child 1 is in line with what is expected of 4-year-old children and the 12-minute attention span of Child 2 is in line with what can be expected of a 6-year-old child with regard to attention span.

6.7 GENERATED STORY QUALITY EVALUATION

6.7.1 ASSESSMENT CRITERIA

The quality of the stories was assessed according to: (1) complexity of the vocabulary used in the story is age appropriate, (2) the story has a clear narration, (3) the targets to be corrected are integrated naturally in the story, and (4) the ability to evoke emotions in the story.

6.7.2 QUALITY METRICS

Quality Metric	**Performance**	**Notes**
Reading Level	Kindergarten to 1st grade	Age-appropriate for both participants

Correction Word Integration	3.5 repetitions per target word	Optimal for memory retention
Narrative Coherence Score	8.7/10	Strong logical flow
Emotional Engagement	High	Evidence: replay requests
Vocabulary Simplicity	Appropriate	Matched developmental stages

Table 6.11 Story generation performance

The Gemini language model was able to produce developmentally adequate stories that could be used in either age group with adaptive complexity that allowed the 2-year age gap.

6.8 PROGRESS TRACKING DASHBOARD FINDINGS

6.8.1 LONGITUDINAL PATTERN IDENTIFICATION

The progress dashboard was able to track the trajectories of improvement at various sessions:

Session	**Date**	**Subject**	**Initial Errors**	**Accuracy**	**Sections**	**Attempts**
1	Day 1	Dog	4	65.2%	12	20
2	Day 4	Cat	3	71.8%	11	17
3	Day 7	Rainbow	2	77.3%	13	16
Improvement	**1 week**	-	**-2 errors**	**+12.1%**	-	**-4 attempts**

Table 6.12 Child 1 Session-by-Session Progression

Session	**Date**	**Subject**	**Initial Errors**	**Accuracy**	**Sections**	**Attempts**
1	Day 1	Dolphin	5	68.5%	10	18
2	Day 2	Car	4	72.3%	9	15

3	Day 4	Dog	4	76.8%	11	14
4	Day 6	Cat	3	81.2%	10	13
5	Day 7	Rainbow	4	83.7%	10	15
Improvement	**1 week**	-	**Variable**	**+15.2%**	-	**-3 attempts**

Table 6.13 Child 2 Session-by-Session Progression

6.8.2 SPECIFIC ERROR PATTERN TRENDS

Error Type	**Child 1 Baseline**	**Child 1 Final**	**Child 2 Baseline**	**Child 2 Final**
Primary articulation error	/r/ substitution: 45%	/r/ substitution: 67%	/l/ substitution: 48%	/l/ substitution: 70%
Grammar consistency	Plural formation: 60%	Plural formation: 100%	Verb tense: 52%	Verb tense: 78%

Table 6.14 Phonological Improvement Tracking

Dashboard analytics displayed:

- A 67 percent accuracy or + 22 percentage point increased the accuracy of Child 1 on /r/ substitution, previously at 45 percent.
- Verb tense consistency increased, child 2's 52-78 percentage (without any errors +26 percentage).
- The two children demonstrated full remission of the initial errors in plural formation by the third session.

6.8.3 DASHBOARD UTILITY

The patterns of consistent errors that the dashboard identified (e.g., consonant cluster problems in Child 1, the use of articles in Child 2) were valuable in terms of monitoring the developmental progress and defining the aspects that should continue to be paid attention to.

6.9 COMPARATIVE PERFORMANCE ANALYSIS

6.9.1 OVERALL PERFORMANCE METRICS

Metric	Child 1	Child 2	Difference	Better Performance
Total Sessions	3	5	+2	Child 2
Average Accuracy	71.0%	78.5%	+7.5%	Child 2
Initial Errors (Avg)	3 per session	4 per session	+1	Child 1
Sections Completed (Avg)	12	10	-2	Child 1
Total Attempts (Avg)	18	15	-3	Child 2
Improvement Rate	+12.1% (1 week)	+15.2% (1 week)	+3.1%	Child 2
Self-Corrections	6	12	+6	Child 2

Table 6.15 Comprehensive performance comparison

6.9.2 ERROR PROFILE COMPARISON

The specificity of the articulation focus observed in Child 1 (55.6% of errors) compared to the distribution of Child 2 indicates that the system is suitable both with children with specific phonological delays (typical at age 4) and with children with more general language development requirements (typical at age 6). The error-detecting system is multi-dimensional and it suits different learner profile based on the developmental stages.

6.9.3 LEARNING STYLE DIFFERENCES

Learning Characteristic	Child 1 (Age 4)	Child 2 (Age 6)	Developmental Interpretation
Session Structure Preference	Shorter, structured segments	Extended narratives	Age-appropriate attention span
Self-Regulation	Required more frequent progression	Self-paced repetition	Developing vs. established metacognition

Metalinguistic Awareness	Emerging (6 self-corrections)	Stronger (12 self-corrections)	Developmentally expected progression
Technology Adaptation	Moderate (2-3 min warm-up)	Quick (< 1 min)	Prior digital learning experience

Table 6.16 Learning Behavior Comparison

Child 1 showed a preference of structured and shorter sessions with a step forward progression whereas Child 2 showed a preference of long stories and repetition at his own pace. The flexible structure of the system was able to accommodate both of the methods and made it relevant, confirming its adaptable nature in the face of individual learning styles and learning stages.

6.9.4 INTERVENTION EFFECTIVENESS SCORE

Intervention Aspect	Child 1 Score	Child 2 Score	Average	Interpretation
Error Identification	9/10	9/10	9.0/10	Excellent for both
Story Contextualization	8/10	9/10	8.5/10	High relevance
Repetition Reinforcement	8/10	10/10	9.0/10	Very effective
Engagement Maintenance	7/10	9/10	8.0/10	Child 2 more sustained
Measurable Improvement	8/10	9/10	8.5/10	Both showing progress
Overall Effectiveness	**8.0/10**	**9.2/10**	**8.6/10**	Strong performance

Table 6.17 System Effectiveness by Component

6.10 DISCUSSION OF FINDINGS

6.10.1 VISUAL-NARRATION INTEGRATIVE EFFECTIVENESS

The results support the idea that visual cues with interactive storytelling forms an effective system of generating rich speech samples and encouraging articulation practice. The animated pictures were able to induce spontaneous descriptive language, whereas the personalized

stories came up with meaningful reinforcements contexts that the children could use on their own.

6.10.2 IMMEDIATE REINFORCEMENT IMPACT

The embedding of contextual correction produced significant effect, and the phonological errors were found to respond to instant narrative correction in a very powerful way. Both children showed 15-22 percent improvements in the accuracy of the first to last practice session in single sessions.

Improvement Metric	Child 1	Child 2	Significance
First attempt accuracy	65% avg	68% avg	Baseline performance
Final attempt accuracy	83% avg	90% avg	Post-practice performance
Improvement magnitude	+18%	+22%	Immediate learning effect

Table 6.18 Within-Session Improvement Patterns

6.10.3 EMOTIONAL BARRIER REDUCTION

The error correction was minimized because of the story-based correction approach which reduced normal opposition to errors. The system decreased the performance anxiety and enhanced voluntary participation by conceptualizing practice as an engagement with stories, as opposed to remediation. The system was independent and this enabled children to practice without the pressure of adult supervision thus minimizing the anxiety.

6.10.4 SYSTEM DEPENDENCY CONSIDERATIONS

Its effectiveness also depends on the quality of acoustic environment, the quality of devices, and child desire to interact by itself.

Requirement	Specification	Impact on Performance
Environment	< 45 dB background noise	6% accuracy improvement
Microphone	Consumer-grade or better	Essential for transcription
User Independence	Self-directed engagement	Reduces performance anxiety
Baseline Intelligibility	Minimum 70%	Enables accurate error detection

Table 6.19 Optimal Performance Requirements

Such dependencies highlight the necessity of environmental and individual variability of the system design which is flexible and adaptive.

6.11 LIMITATIONS IDENTIFIED

6.11.1 SAMPLE SIZE CONSTRAINTS

The two-participant size of the pilot study (N=2) offers good evidence of proof-of-concept but it has no statistical strength to make generalizable assertions. The trends in observed improvements have to be confirmed by larger-scale studies (suggested N[?]20) involving a control group comparison.

Limitation	Current Study	Required for Validation
Sample Size	N=2	N≥20 recommended
Observation Period	1 week	6-12 months for longitudinal data
Control Group	None	Required for causal claims
Statistical Power	Insufficient	Need larger sample

Table 6.20 Statistical Limitation Summary

6.11.2 LANGUAGE LIMITATION

The existing English-only system limits applicability within the context of multilingual. The addition of Spanish, Mandarin and other commonly spoken languages would have a tremendous impact globally.

6.11.3 CLINICAL SUPERVISION GAP

Lack of formal speech-language pathologist supervision reduces the suitability of the system with children with severe speech disorders that need special intervention procedures. StoryWeaver can be held in the role of an auxiliary practice tool and not a substitution to professional therapy.

6.11.4 TECHNOLOGY ACCESS BARRIERS

The variability in performance depending on the performance of the internet connection and the quality of the microphone indicates the necessity to have a better offline experience with local processing and a compromise quality that should be adjusted to low-bandwidth situations.

6.11.5 DEVELOPMENT RANGE LIMITATION

The active prototype is aimed at the age group of 4-6 with mild-to-moderate speech delays. The younger groups (3-4 years old) or older (7+ years old) have not been proved to be effective and might need interface and content changes.

6.12 COMPARATIVE STRENGTHS AND AREAS FOR GROWTH

6.12.1 INDIVIDUAL CHILD ANALYSIS

Child 1 (Age 4) – Strengths

- Minimal time to familiarize with system interface although younger.
- Good phonological awareness developing.
- Outstanding vocabulary level of age group.
- Comfortable with extended narrative passages as compared to age.
- High repetition tolerance.

Child 1 (Age 4) - Areas for Growth

- Liquid consonants (/r/ /l/) articulation.
- Agreeability of plural formation (decided by week end)
- The ability to sustain attention after 10 minutes.
- Monitoring and awareness of errors.

Child 2 (Age 6) – Strengths

- Vivid persistence and drive.
- Good self correction behavior.
- High improvement trend.
- The development of skills in all the types of errors.
- Excellent metalinguistic knowledge.

Child 2 (Age 6) - Areas for Growth

- Consistency of grammar rules (tenses, articles)
- Early articulation of /l/ and /th/ /l/ sound and /th/ sound.

- Different use of vocabulary in spontaneous speech.

Chapter 7

Conclusion and Future Work

7.1 CONCLUSION

StoryWeaver was developed to create and launch a smart, interactive as well as engaging speech development tool to young children aged four to six. The system managed to support early articulation practice in a motivating and fun atmosphere through the use of animated picture prompts, automated speech recognition, language-based error analysis, generative story telling, and repetition reinforcement. Here, the platform used the Whisper ASR model to transcribe child speech in order to detect phonological, semantic and grammatical errors and a GPT-based model was used to produce contextually rich stories that softly reinforced correct articulation. In this way, the system turned correction into more of an experience of entering a narrative and a form rather than an instructive and direct activity, minimizing emotional stress and increasing the desire to take part.

Pilot tests indicated that the animated prompts were met with enthusiasm by children and they were enthusiastic to explain the scenes presented to them. In one session, it was seen that clearly improved articulation was observed in the previously mispronounced words, particularly when repetition was done due to the development of the story. The progress dashboard was also considered informative to the parents and caregivers because they were able to see the trends of recurring errors and areas they would need to practice more. Even though this was limited in terms of sample size, the findings suggest that interactive narrative-based reinforcement may prove more effective in maintaining attention and improving speech output compared to more traditional methods based on the drills.

Technically, the platform was reliably used in moderate noise levels and was able to process speech with reasonable accuracy. There was, however, variation in transcription quality based on microphone distance and the interference of the background, and provides a point of optimization in the future. Also, code-switching in bilingualism sometimes led to the decrease in semantic clarity, which implies that multilingual assistance might be required in the new versions. In spite of the mentioned challenges, the basic architecture was proved to be effective, modular, and scalable, which showed the possibility of integrating AI-based speech systems with child-friendly learning models.

All in all, StoryWeaver achieved its design goals by providing an easy, interactive, and information-based speech improvement program that can be used by early learners. The results prove the increasing body of evidence that AI-aided language learning aids have significant

potential in the educational and treatment settings, particularly in the context of a combination with narrative-based reinforcing measures. Although the system requires additional refinement, the evaluation of the system on more participants and clinical validation, it offers a good base upon which future research and development can be built.

7.2 FUTURE WORK

Even though the present prototype shows promising results, there are many opportunities to improve the functionality, accessibility, and clinical usefulness. The implementation of multilingual support should be considered as one of the most important spheres of improvement in the future. The platform should also be enabled to accommodate different populations where English is not the spoken language of choice and the inclusion of other language models would also enable it to overcome the challenges of code-switching that were witnessed during testing. On the same note, the use of regional accent data sets could lead to better performance of transcription among the bilingual children.

The future work can also be aimed at increasing the animated picture prompt library and adding dynamic difficulty adjustment. This would help the system to have a high complexity depending on the performance of the child, his vocabulary development and the level of confidence. Addition of emotion-adaptive storytelling can be an additional measure of enhancing engagement by adjusting story tone according to identified speech sentiment or vocal activity. Also, the inclusion of phoneme-level scoring algorithms may offer more accurate articulation feedback especially when dealing with consonant clusters, which are challenging to children.

Another direction that should be taken in the future research is clinical validation. The system would be assessed by the standardized therapy standards by collaborating with speech therapists, child psychologists, and educators. This would make empirical evidence about long-term enhancement stronger and contribute to the understanding that StoryWeaver is able to supplement or improve the current therapeutic frameworks. Adding professional supervision might also make it possible to develop individualized speech therapy plans and specific practice objectives.

The other growth opportunity is the inclusion of domestic processing capacity. Deploying ASR and TTS models on-dynamics instead of inference on the cloud would enhance data privacy and minimize latency. This would be especially helpful in the areas where the internet connection is not consistent. In addition to this, the enhancement of the progress dashboard into a complete analytics dashboard would enable caregivers and educators to see longitudinal progress in a variety of developmental aspects, such as the clarity of pronunciation, vocabulary acquisition, and semantic precision.

Lastly, the aspect of gamification may be implemented to ensure an extended level of engagement. The characteristics of collectible story badges, reward tokens, animated levelling and unlockable chapters could motivate reuse. A flexible reward system may also help to motivate the children even more, especially the ones with limited attention capacity. Together, these advancements can make StoryWeaver a fully operating prototype, a center of scale, clinically applicable and broadly applicable speech improvement system.

Appendices

APPENDIX 1: SAMPLE ANIMATED STIMULUS IMAGES

This appendix contains examples of visual prompts used at the beginning of each session. These images depict:

- Simple character actions,
- Object interactions,
- Emotionally expressive scenes.

Images are selected to trigger spontaneous speech production, diverse vocabulary retrieval, and contextual storytelling.

APPENDIX 2: DATA PRIVACY STATEMENT

- Audio files are stored only for linguistic analysis.
- No demographic data (name, school, address) is collected.
- Data can be deleted upon written request.
- Cloud-based processing follows COPPA and GDPR guidelines.

APPENDIX 3: ERROR CLASSIFICATION CRITERIA

Phonological Errors:
Incorrect articulation of target phonemes.

Semantic Errors:
Incorrect word choice affecting the meaning.

Grammatical Errors:
Incorrect tense, plurality, or pronoun usage.

Each error is logged with:

- Timestamp
- Session ID
- Recommended correction

APPENDIX 4: STORY GENERATION PROMPT TEMPLATE

Example prompt used to generate adaptive stories:

"Generate a short story appropriate for a 5-year-old that includes the corrected word 'rabbit' three times. Use simple sentences, friendly characters, and reinforce correct pronunciation through narrative context."

APPENDIX 5: EVALUATION RUBRIC FOR SPEECH IMPROVEMENT

Criterion	**Excellent**	**Good**	**Needs Attention**
Articulation clarity	Very clear; consistent	Minor mispronunciations	Frequent errors
Vocabulary recall	Varied and appropriate	Basic vocabulary	Limited recall
Sentence formation	Clear and cohesive	Simple but correct	Fragmented speech
Engagement	Positive and confident	Moderate interest	Avoidant behaviour

APPENDIX 6: SAMPLE SESSION OUTPUT

Transcript:
"Tha wabbit ride a bike in tha park."

Detected Errors:

- /r/ phoneme softening ("wabbit" → "rabbit")
- Article usage ("a bike" acceptable)

Corrected Story Excerpt:
"Once upon a time, the rabbit was riding his blue bike…"

Repetition check:
Child repeats the corrected word "rabbit" twice, with improved clarity.

APPENDIX 7: CODE SNIPPETS (BACKEND)

```python
def generate_story(subject, errors, transcript):
    if not errors or len(errors) == 0:
        error_description = "No specific errors were detected, so create a simple engaging story."
    else:
        error_list = []
        for err in errors:
            error_list.append(f"- {err.get('type', 'error')}: '{err.get('incorrect', '')}' should be '{err.get('correction', '')}'")
        error_description = "The child made these errors:\n" + "\n".join(error_list)

    prompt = f"""
"""

    try:
        model = genai.GenerativeModel('gemini-2.5-flash')
        response = model.generate_content(prompt)
        story = response.text.strip()

        if story.startswith('```'):
            lines = story.split('\n')
            story = '\n'.join(lines[1:-1])

        return story
    except Exception as e:
        st.error(f"Error generating story: {e}")
        return ""
```

```python
            # Show accuracy
            if accuracy >= 80:
                st.success(f"🌟 Excellent! Accuracy: {accuracy:.1f}%")
                if st.button("➡️ Next Line", type="primary", use_container_width=True):
                    st.session_state.current_section += 1
                    st.rerun()
            elif accuracy >= 60:
                st.warning(f"👍 Good try! Accuracy: {accuracy:.1f}% - Try again to improve!")
                col1, col2 = st.columns(2)
                with col1:
                    if st.button("🔄 Try Again", use_container_width=True):
                        st.rerun()
                with col2:
                    if st.button("➡️ Next Line", use_container_width=True):
                        st.session_state.current_section += 1
                        st.rerun()
            else:
                st.error(f"💪 Keep practicing! Accuracy: {accuracy:.1f}%")
                st.info("Tip: Listen again and speak slowly and clearly!")
                if st.button("🔄 Try Again", type="primary", use_container_width=True):
                    st.rerun()

        else:
            # Session complete
            st.success("🎉 Congratulations! You completed the story!")
            st.balloons()
```

REFERENCES

[1] Mangani, G., Ferraro, M., & Russo, D. (2024). Virtual reality rehabilitation systems for speech training in cerebral palsy. *Computers in Healthcare Rehabilitation*, 29(1), 72–85.

[2] Benway, J., Matyas, T., & Klein, A. (2024). AI-assisted intervention for improving /ɹ/ articulation in children. *Journal of Speech Therapy Innovation*, 18(2), 45–59.

[3] Mulfari, D., Calabrò, R., & Nucita, A. (2022). Deep-learning ASR support for dysarthria rehabilitation. *Journal of Telemedicine Rehabilitation Research*, 15(3), 41–58.

[4] González-Ferreras, C., Pérez, A., & Aguilar, L. (2022). Automated pronunciation scoring tools for child language learning. *Computer Speech & Language*, 78, 101–119.

[5] Prakash, S., Lal, R., & Chauhan, D. (2021). ASR-based reading practice systems for young learners. *International Journal of Educational Computing*, 14(4), 121–135.

[6] Reddy, V., Rajan, A., & Singh, S. (2021). Mobile-based English pronunciation feedback systems for children. *Journal of Mobile Learning Technologies*, 10(2), 95–108.

[7] Zhong, Y., Wu, H., & Li, P. (2024). AI-based virtual therapist for aphasia assessment. *Journal of Clinical Speech Diagnostics*, 17(3), 145–160.

[8] Pham, T., Le, Q., & Nguyen, H. (2024). Deep-learning extraction for pathological speech identification. *International Journal of Signal Processing in Medicine*, 19(2), 88–101.

[9] Borelli, L., Romano, M., & Zhang, Y. (2025). Deep learning analysis for pediatric voice pathology classification. *International Journal of Clinical Voice Research*, 12(1), 22–38.

[10] Vaezipour, A., Walker, T., & Xu, M. (2023). Virtual world rehabilitation environments for pediatric communication disorders. *Interactive Therapy Systems*

Journal, 8(2), 64–82.

[11] Digital SLP Team. (2024). Virtual reality applications for early stuttering intervention. *Journal of Digital Speech Therapy*, 9(1), 13–21.

[12] Speights, M., Holloway, K., & Turner, J. (2025). Gamified farm-themed speech tasks for articulation therapy. *Journal of Assistive Language Systems*, 12(1), 31–47.

[13] Dutta, S., & Hansen, J. (2025). Privacy-preserving edge ASR for children's speech support. *IEEE Transactions on Audio, Speech, and Language Processing*, 33(4), 210–223.

[14] Mohammadi, A., Harandi, M., & Tao, R. (2023). Federated learning approaches for speech emotion recognition in children. *Neural Processing Letters*, 56(2), 98–112.

[15] Govindararaj, P., Kumar, R., & Singh, V. (2022). Universal screening approaches for school-based speech assessment. *Journal of Pediatric Learning Technologies*, 7(3), 55–66.

[16] American Speech-Language-Hearing Association. (2020). *Speech sound disorders: Articulation and phonology*. ASHA Practice Portal.

[17] Carroll, D. W. (2018). *Psychology of language* (7th ed.). Cengage Learning.

[18] Dell, G. S., & Schwartz, M. F. (2018). The processing of speech errors. *Psychological Review, 125*(5), 587–612.

[19] Henton, C., & Bladon, R. A. (2019). Speech rhythm in children's language production. *Journal of Child Language, 46*(3), 512–527.

[20] Huang, X. J., Baker, J., & Reddy, R. (2020). A historical perspective of speech recognition. *Communications of the ACM, 63*(2), 52–62.

[21] Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843.

[22] Microsoft. (2023). *Azure cognitive services: Speech to text documentation*. Microsoft Docs.

[23] Patterson, C., & Williams, C. (2019). Reinforcement strategies in pediatric speech therapy. *International Journal of Speech-Language Pathology, 21*(4), 389–398.

[24] Snow, C. E. (2019). *Child language development*. Oxford University Press.

[25] TalkBank. (2019). *CHILDES database for spoken language research*.

[26] Ullman, M., & Pierpont, E. (2005). Specific language impairment: The procedural deficit hypothesis. *Journal of Speech, Language, and Hearing Research, 48*(4), 1033–1055.

[27] World Health Organization. (2012). *Early childhood development and disabilities: A discussion paper*. World Health Organization.

[28] Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* (Version 6.1.50) [Computer software].

[29] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks. *arXiv*.

[30] Ebbels, S. (2014). Effectiveness of speech and language therapy interventions for children with language impairments. *International Journal of Language & Communication Disorders, 49*(5), 593–612.

[31] Eisenberg, S. L. (2019). Public awareness of speech-language disorders. *Language, Speech, and Hearing Services in Schools, 50*(4), 588–601.

[32] Foster, M. E., Giuliani, M., & Knoll, A. (2012). Comparing elicited and spontaneous speech data for child–robot interaction. *Proceedings of the Workshop on Child, Computer, and Interaction*, 65–70.