Abstract

This project introduces an AI-powered web application designed to address speech disorders in children and emotional challenges in adults by providing an accessible, interactive, and scalable alternative to traditional therapy. Traditional speech and emotional therapy methods are costly, inaccessible, and lack personalized engagement. The platform features a real-time AI avatar that guides children through interactive speech exercises, providing corrective feedback and structured practice to improve pronunciation and fluency. For adults, the system leverages AI-driven emotional support, analyzing voice tone and text input to deliver personalized, calming responses. The application enhances user engagement through gamified learning, progress tracking, and adaptive AI features, ensuring a personalized and effective experience. Guardians can customize therapy sessions and monitor progress through an intuitive dashboard. By integrating speech recognition, and deep learning models, this software overcomes the limitations of traditional therapy, offering a scalable, data-driven, and user-friendly solution for individuals needing speech training and emotional well-being support.

Table of Contents

[1. Introduction 1](#_Toc28375)

[2. Problem Description 1](#_Toc28376)

[2.1. Primary Scope 1](#_Toc28377)

[2.2. Final Deliverable of the Project and Beneficiaries 2](#_Toc28378)

[2.3. Objectives 2](#_Toc28379)

[2.4. Novelty 2](#_Toc28380)

[2.5. Literature Review 3](#_Toc28381)

[2.6. Fluenti's Unique Approach 3](#_Toc28382)

[3. Methodology 3](#_Toc28383)

[3.1. Dataset & Training Data 4](#_Toc28384)

[4. Feasibility Plan 6](#_Toc28385)

[4.1. Resource Requirement 7](#_Toc28386)

[4.1.1. Expertise of the Team 8](#_Toc28387)

[4.1.2. Tools / Technology 8](#_Toc28388)

[4.1.3. Budget 8](#_Toc28389)

[4.2. Risks Involved 8](#_Toc28390)

[5. Key Milestones and Schedule 9](#_Toc28391)

[5.1. Key Milestones 9](#_Toc28392)

[5.2. Gantt Chart 11](#_Toc28393)

[REFERENCES 13](#_Toc28394)

# Introduction

Speech disorders in children and emotional challenges in adults are significant barriers to effective communication, yet they often remain unaddressed due to high therapy costs, limited access to professionals, and lack of awareness. In underserved areas, the shortage of speech therapists and mental health experts makes consistent support difficult, leaving many individuals without the help they need [1].

This project proposes an AI-powered web application to bridge this gap. For children, the platform features a real-time AI avatar that guides speech exercises, provides corrective feedback, and engages in interactive learning. For adults, the system offers personalized emotional support through conversational AI, tailored to individual needs [2]. The application also includes progress tracking, gamified features, and adaptive learning to ensure an engaging and effective experience [3]. Guardians can customize sessions and monitor improvements, making speech training and emotional assistance more accessible and scalable. By leveraging AI, this solution addresses the limitations of traditional therapy methods, offering a comprehensive and user-friendly approach for diverse users.

# Problem Description

## Primary Scope

The web application addresses two key areas: speech therapy for children and emotional support for adults, with an AI avatar capable of interacting in both English and Urdu. For children with speech disorders, an AI-powered avatar will speak words or sentences in either language, guiding the child to repeat them with real-time feedback on pronunciation accuracy. Guardians can customize exercises based on the child’s needs, ensuring a personalized learning experience. For adults, an AI chatbot will analyze voice tones in both English and Urdu, detecting emotional states like stress or anxiety and offering supportive, calming responses accordingly. The platform seamlessly integrates bilingual AI-driven speech therapy and emotional support, enhancing accessibility, engagement, and effectiveness for diverse linguistic users.

## Final Deliverable of the Project and Beneficiaries

Final Product: A web-based AI-driven platform providing speech therapy and emotional support.

**Beneficiaries:**

1. Adults looking for emotional support and stress relief.
2. Children with speech disorders.
3. Guardians monitoring speech progress.

## Objectives

Here are the key objectives for the Conversational AI Chatbot project:

1. Develop a secure AI-powered platform for speech therapy and mental health support.
2. Implement an interactive AI avatar to assist users.
3. Provide personalized, gamified speech therapy exercises.
4. Offer real-time emotional support through AI-driven responses.
5. Enable progress tracking dashboards for Guardians

## Novelty

This project combines AI-driven emotional support, a real-time interactive AI avatar, and personalized speech therapy into a single platform. This software features a dynamic avatar that speaks, listens, and reacts, offering engaging experiences. For children, gamified speech exercises with real-time feedback make learning fun and effective [3], while adults benefit from voice-based emotion detection and tailored support. User-friendly dashboards for guardians enhance accessibility. By merging these features, the platform offers a unique solution for speech improvement and emotional support, setting it apart from conventional chatbots and therapy tools.

## Literature Review

The use of AI in speech therapy and emotional support has seen rapid advancements, leading to the development of various applications. While existing solutions provide valuable assistance, they have notable limitations that Fluenti aims to address. The summary of comparison of the proposed Flunti with other similar applications are discussed in table 2.4.1.

Table 2.4.1 provides an overview of existing applications that are similar to Fluenti.

|  |  |  |
| --- | --- | --- |
| **Application** | **Features** | **Limitations** |
| Replika [4] | Conversational AI for emotional support. | Lacks speech therapy capabilities. |
| Woebot [5] | Uses CBT-based mental health strategies. | No voice-based interactions. Does not include an AI avatar. |
| Calmi.co [6] | AI-driven mental wellness chatbot with emotion analysis. | Lacks interactive speech therapy features. Does not include an AI avatar. |

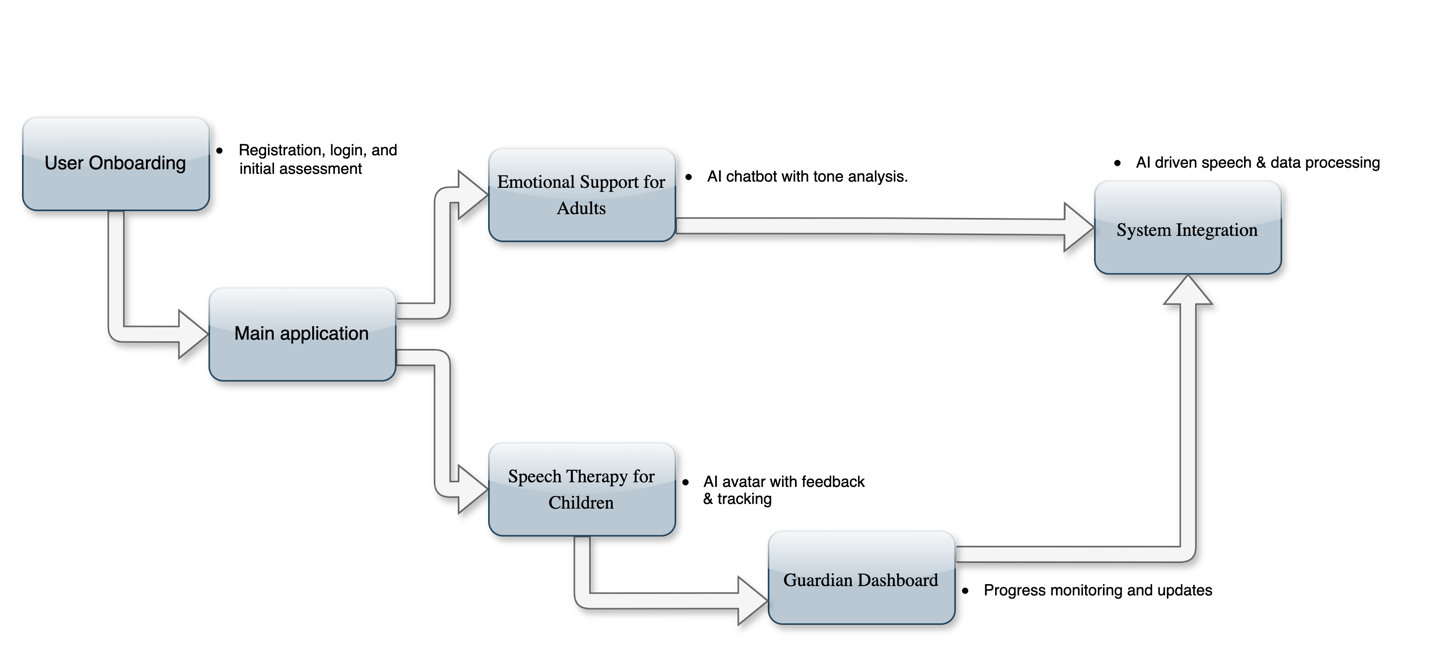
## Fluenti's Unique Approach

Fluenti combines AI-driven speech therapy and emotional support in one platform. It supports bilingual interaction (English & Urdu), real-time pronunciation feedback, gamified learning, an interactive 3D AI avatar, and an initial assessment system that suggests personalized exercises based on a child's speech abilities. Unlike existing solutions, Fluenti offers a unified, accessible AI system for all users.

# Methodology

As shown in Figure 3.1, the development process starts with Requirement Analysis, identifying the needs of guardians and individuals seeking mental health support. System Design focuses on building the architecture, API structure, and database models to ensure seamless integration. AI & Avatar Development involves implementing speech recognition, NLP, and a 3D animated AI avatar for real-time interaction. The system conducts an initial speech assessment to analyze pronunciation and fluency, then suggests personalized exercises. The AI adapts therapy as the user progresses. Frontend & Backend Development ensures a user-friendly UI (React next js), a robust backend (Node.js + Express.js), and a scalable database (MongoDB). Integration & Testing validates speech accuracy, emotional response effectiveness, and overall UI functionality. Deployment & Evaluation hosts the application on Firebase and AWS, ensuring scalability, continuous monitoring, and AI model updates for optimal performance.

An Agile approach ensures continuous testing, AI model refinement, and optimization of speech recognition and emotional analysis. This flexible, collaborative process adapts to diverse speech patterns and emotions, ensuring a high-quality, scalable system in English and Urdu.

 ***Figure 3.1: User Experience Flow Diagram***

## Dataset & Training Data

Fluenti utilizes diverse datasets to enhance speech recognition and emotional analysis:

1. Mozilla Common Voice (Urdu Dataset)

The Mozilla Common Voice Urdu Dataset is a crowdsourced speech dataset designed to improve Automatic Speech Recognition (ASR) systems for the Urdu language.

• Features:

* Contains diverse Urdu accents and dialects, enabling better recognition across regions. o Includes text-audio alignment, facilitating accurate speech-to-text transcription. o Supports speaker diarization, allowing differentiation between multiple speakers.
* Open-source and regularly updated, making it adaptable for custom speech therapy applications.

2. IEMOCAP (Interactive Emotional Dyadic Motion Capture Database)

The IEMOCAP dataset is widely used for emotion recognition in speech and is essential for detecting stress, anxiety, and emotional variations.

• Features:

* A multi-modal dataset containing audio, text transcriptions, and video.
* Includes acted and spontaneous emotional dialogues, capturing real-world emotional variations.
* Covers 10 distinct emotions, allowing precise emotional classification.
* Enables speech-based emotion detection, useful for mental health monitoring and AIdriven conversational therapy.

3. TIMIT Corpus

The TIMIT Corpus provides high-quality speech recordings with phoneme transcriptions, making it essential for pronunciation correction and phonetic analysis.

• Features:

* Contains phoneme-level transcriptions for 630 speakers from 8 major US dialects.
* Includes high-quality, time-aligned speech recordings, useful for speech synthesis and ASR training.
* Supports accent adaptation, allowing personalized pronunciation feedback.
* Facilitates linguistic research by providing detailed phonetic structures.

4. GoEmotions

The GoEmotions dataset is designed for sentiment analysis and emotion classification, enhancing chatbot emotional intelligence.

• Features:

* Covers 27 nuanced emotions, including admiration, gratitude, and fear. o Consists of human-annotated data, ensuring high accuracy in emotion recognition.
* Supports multi-label classification, allowing detection of multiple emotions within a single sentence. o Enables context-aware sentiment detection, improving chatbot and AI responsiveness.

These datasets enable Fluenti to deliver accurate speech therapy and emotionally responsive interactions.

# Feasibility Plan

Resource Requirements

1. Frontend**:** React next js + readyplayer (for 3D Avatar)
2. Backend**:** Node.js + Express.js
3. Database**:** MongoDB (NoSQL) + Firebase
4. AITools**:** Google Speech-to-Text, Azure Speech Services, TensorFlow.js
5. Gamification**:** Phaser.js for interactive speech exercises

Risks & Mitigation

1. Speech Recognition Accuracy: Use Google/Azure APIs for better detection.
2. User Adoption: Keep UI minimal & user-friendly.
3. DataPrivacy: Ensure GDPR & COPPA compliance for children’s data.

## Resource Requirement

1. Software Development Tools:

Integrated Development Environment (IDE): Visual Studio Code for coding and development.

Web Development Frameworks: React next js for the frontend, Node.js & Express.js for the backend.

Database Management System: MongoDB (NoSQL) for storing user profiles, speech progress data, and chatbot interactions.

1. Human Resources:

Project Team Members: Frontend developer, backend developer, database administrator, and UI/UX designer.

AI Engineers: To develop speech recognition models and chatbot interactions.

Quality Assurance (QA) Testers: For testing speech accuracy, AI responsiveness, and overall app performance.

Speech Therapists & Mental Health Experts: To ensure clinical accuracy of exercises and chatbot responses.

1. AI & Speech Recognition Tools:

OpenAI GPT API for conversational AI in mental health support.

Google Speech-to-Text API, Vosk API, or Azure Speech Services for pronunciation analysis [7].

Implement speech recognition, NLP, and 3D animated AI avatar using Ready Player Me [8].

1. Testing and Deployment Tools:

Automated Testing Frameworks: Selenium for frontend testing, Postman for API testing.

Continuous Integration/Deployment (CI/CD): GitHub Actions for automated deployment.

Cloud Hosting: Firebase/AWS for real-time speech processing and chatbot responses.

### 4.1.1.1. Expertise of the Team

Team members have foundational knowledge in web development, AI, and database management. Their experience and enthusiasm for AI-driven applications ensure dedicated effort towards the project's successful completion.

### 4.1.1.2. Tools / Technology

The project will utilize open-source tools and technologies, including:

* Development: Visual Studio Code, GitHub
* Frontend: React next js
* Backend: Node.js, Express.js
* Database: MongoDB
* AI Tools: Google Speech-to-Text, Azure Speech API, OpenAI for chatbot
* Avatar Animation: Ready Player Me for real-time AI avatar interactions.

## Risks Involved

* Speech Recognition Accuracy Issues: AI may struggle with diverse accents or severe speech disorders, mitigated by training on diverse datasets.
* Integration Complexity: Combining multiple AI technologies may cause technical challenges, managed via modular design and iterative testing.
* User Adoption Challenges:Users may find the platform initially challenging to use, eased by an intuitive UI and comprehensive user guides.
* Scalability Issues: High user demand may strain the platform, mitigated by using scalable cloud infrastructure and load balancing techniques.

# Key Milestones and Schedule

## Key Milestones

Table 5.1 outlines the major milestones of the project, including their descriptions and estimation. ***Table 5-1: Breakdown of work in form of milestones***

|  |  |  |  |
| --- | --- | --- | --- |
| S. No. | S.No. of  Predecessor  Milestone | Key Milestone Name / Description | Duration (personhours1) |
| 1 | - | Requirement Gathering: Conduct surveys to gather project requirements. | 30 |
| 2 | 1 | System Design: Develop system architecture, database design, and UI wireframes. | 70 |
| 3 | 2 | AI & Avatar Development: Implement Ready Player Me avatar and integrate AI-powered speech therapy. | 100 |
| 4 | 3 | Frontend Development: Build the user interface using React next js and TailwindCSS**.** | 100 |
| 5 | 4 | Backend Development: Implement server-side logic and database interactions using Node.js and Express.js. | 100 |

|  |  |  |  |
| --- | --- | --- | --- |
| 6 | 5 | Integration of Frontend and Backend:Connect frontend and backend components and ensure proper communication. | 30 |
| 7 | 6 | Testing: Perform unit testing, integration testing, and user acceptance testing (UAT). | 60 |
| 8 | 7 | Deployment: Set up hosting, deploy the application, and register the domain name. | 20 |
| 9 | 8 | Training and Documentation: Create user manuals and provide training sessions for administrators and users. | 20 |
| 10 | 9 | Final Evaluation and Reporting**:** Analyze performance metrics, gather feedback, and compile a final project report. | 20 |

# REFERENCES

1. Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.
2. Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2023). Human--AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. Nature Machine Intelligence*,* 5(1), 46-57.
3. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining" gamification". In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments (pp. 9-15).
4. “Replika.” replika.com. Accessed March 12, 2025. https://replika.com/. [5] Woebot Health, March 3, 2025. https://woebothealth.com/.
5. Calmi. Accessed March 12, 2025. https://calmi.so/.
6. Bumann, N. (2023). Automated Chatbot Using Speech-to-Text and Text-to-Speech with Mobile App Integration. Journal of Artificial Intelligence and Applications*,* 12(3), 45-60.
7. Altundas, S., & Karaarslan, E. (2023). Cross-platform and personalized avatars in the metaverse: Ready player me case. In Digital twin driven intelligent systems and emerging metaverse (pp. 317-330). Singapore: Springer Nature Singapore.