Development of a Generative AI Application for Therapy of Speech-Impaired Patients

Muhammad Tayyab, Mohammad Hashir Sajjad, Mohammad Uzair, Arbab Masood Ahmad, Syed Ghulam Moeen Ud Din Shah Banoori

Department of Computer Systems Engineering, University of Engineering and Technology, Peshawar, Pakistan

***Correspondence:** 22tayyabkhan22@gmail.com, hashirsajjad09@gmail.com, uzairsiddiq@gmail.com, arbabmasood@uetpeshawar.edu.pk, banoorimoeen@gmail.com

Citation | Tayyab. M, Hashir. M, Uzair. M, Ahmad A. M, Banoori. S. G. M. U. D. S, "Development of a Generative AI Application for Therapy of Speech-Impaired Patients", IJIST, Vol. 07, Issue. 02 pp 1334-1340, June 2025

Received| June 04, 2025 **Revised|** June 25, 2025 **Accepted|** June 27, 2025 **Published|** June 30, 2025.

With the use of this web application, speech therapy can be provided to individuals with speech disabilities through an easy-to-use and engaging interface, which allows them to receive treatment from anywhere. Conventional treatment often requires the patient's presence, which can be hard due to their location, finances, or even other health conditions. The application helps users practice articulation, fluency, and pronunciation with the help of guided exercises, visual aids, and voice recognition technology. Users are provided with a personalized treatment plan that reviews their progress over time and provides them with feedback, reports, and reminders to help ensure users are motivated and consistent. This new technology will increase the availability of services to patients in underserved regions, which will improve their communication as well as the quality of life.

Keywords: Gen AI, Speech Therapy, Large Language Models

Introduction:

The disorders of dysarthria and stuttering significantly hinder individuals' ability to communicate, often leading to challenges in social, personal, and professional life. These issues may arise from neurological diseases, brain trauma, or developmental disorders, affecting self-esteem and quality of life. Traditional speech therapy remains the most widely adopted treatment. It involves one-on-one sessions with speech-language pathologists, using structured exercises to enhance articulation and fluency. While effective, such methods are costly, time-consuming, and inaccessible to those living in rural or underserved regions.

Recent advances in artificial intelligence (AI), especially in generative AI and natural language processing (NLP), offer promising alternatives for augmenting traditional speech therapy. By leveraging these technologies, remote and personalized treatment can be provided to individuals who might otherwise lack access to professional care.

Despite these developments, there remains a gap in practical, scalable implementations of AI-driven speech therapy solutions designed specifically for people with speech impairments.

Background and Related Work:

Speech impediments pose challenges to millions of people around the globe [1][2], having a huge impact on their communication and engagement in social, educational, and occupational endeavors. They result from multiple causes, such as neurological conditions like strokes [3] and disorders such as autism spectrum disorder [4], or even physical traumas that damage the vocal cords and other parts of the brain associated with speech [5]. Most individuals have relied on speech therapy in the past, where guidance is provided [6] to enhance clarity, fluency, and assurance in the person's spoken language. Unfortunately, traditional speech therapy comes with a host of challenges [7]. Often, it is costly due to the session costs compounding over time, and it depends on the availability of qualified speech-language pathologists (SLPs), who may be limited in number in rural or underserved regions. Additionally, traditional therapy usually requires a lot of face-to-face contact, which is not accommodating for patients who may have issues with transportation, busy schedules, or other health restrictions. This lack of flexibility means that many people can remain unsupported in addressing their speech issues. The Artificial Intelligence (AI) revolution is poised to solve problems like these [8], however. The development of speech therapy applications powered by AI technologies can implement cost-effective, accurate, and personalized therapy techniques. This research proposes the integration of an AI Chatbot with a web application. Users will be able to practice and improve their speech through personalized exercises, immediate feedback, and progress tracking, all from the comfort of their homes. This approach also reduces the total costs for people who do not have easy access to professional care by minimizing the number of required in-person sessions. The main aim of this approach is to utilize artificial intelligence to provide professional, affordable, scalable, and accessible speech therapy services to everyone. This groundbreaking platform seeks to improve the quality of life for people with speech disabilities by empowering them to communicate better and more confidently, by bridging the gaps of traditional models of therapy.

Objectives:

The objective of this research is to develop and evaluate an AI-powered web application for delivering remote, personalized speech therapy. This system integrates automatic speech recognition (ASR), real-time generative feedback, progress tracking, and gamified learning to enhance user engagement and outcomes. Unlike existing tools that offer static exercises or limited feedback, this solution uses NLP and AI to provide dynamic, tailored therapy for individuals with mild to moderate speech impairments.

Novelty Statement:

This research aims to fill the gap by developing and evaluating a web-based AI application that enables remote, personalized speech therapy using speech recognition, synthesis, and adaptive learning techniques. The novelty lies in its combination of speech analysis, feedback generation, and accessibility, enabling therapy from home without the constant presence of a human therapist.

Methodology:

This research adopted an applied development approach grounded in real-world user needs and informed by speech therapy best practices. The development process followed Agile methodology with iterative sprints, daily stand-ups, and continuous integration to ensure progressive refinement of features based on user testing.

The system architecture consists of four core components:

Frontend Interface: Built with React.js, the interface enables users to interact with the therapy modules, receive real-time AI feedback, and view progress reports.

Backend Server: Developed in Node.js with Express, this component handles user authentication, session management, and API connections with AI services.

AI/NLP Engine: Amazon Lex was used to process speech inputs and generate personalized feedback using natural language processing techniques.

Database: MongoDB is used for storing user profiles, progress logs, session history, and system configurations.

Key technologies also include JWT for secure login, Google Speech-to-Text for ASR (automatic speech recognition), and hosting services like AWS and Firebase for deployment.

The system is composed of five main modules:

1. User Management (registration, login, role-based access).
2. Speech Therapy Interface (task display and speech recording).
3. AI Feedback Engine (real-time pronunciation and fluency evaluation).
4. Progress Tracker (charts and performance logs).
5. Gamification Layer (badges, points, and levels to boost motivation).

Testing was performed at multiple levels: unit, integration, usability, and performance. Tools like Postman, Jest, and Lighthouse were used to verify functionality and response time. Ethical considerations were also addressed, including informed consent and compliance with data privacy standards such as GDPR and HIPAA.

System Architecture:

The proposed system comprises three key components:

Speech Recognition Module: Converts user speech to text using an ASR model [9]. For this, we employed advanced ASR engines such as Whisper and Deep Speech, which are particularly effective in handling speech variability through model fine-tuning. **Generative AI Engine:** Processes input and generates adaptive therapy exercises [8]. **Speech Synthesis Module:** Provides real-time auditory feedback using deep learning models like GTTS [10]. GTTs were integrated for initial text-to-speech output due to their simplicity. However, future integration may involve Tacotron or Fast Speech models for more lifelike vocal feedback.

Data Collection and Training:

The system is trained on diverse speech datasets, including pathological speech samples. Transfer learning is applied to fine-tune models for speech-impaired patterns.

User Interaction Model:

Patients engage with the system via a user-friendly web application. The AI adapts exercises based on real-time progress and feedback.

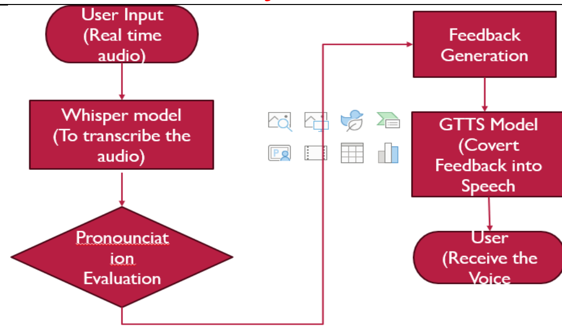

Figure 1. Methodology Flowchart

Results and Discussion:

The application was evaluated through functional validation, usability trials, and speech recognition tests. Functional Testing confirmed that all core modules were operational. The AI feedback system worked reliably, though minor delays were noted under weak internet conditions. Usability Testing involved 12 participants, including individuals with speech impairments and speech therapy students. Participants rated the app highly for ease of use, engagement, and clarity of feedback.

Table 1. Usability Testing Data

Criteria	Average Rating (1–5)
Ease of Navigation	4.6
Real-TimeFeedback Clarity	4.4
Customization Options	4.2
Engagement (Gamification)	4.7
Overall Satisfaction	4.5

In the Speech Recognition Test, 30 phrases were evaluated across users with varying speech clarity:

Table 2. Speech Recognition Test Data

User Group	Accuracy (%)
Non-impaired users	93.5%
Mild impairment	85.2%
Moderate impairment	76.1%

The results matched industry expectations for mild impairment but highlighted challenges with more severe conditions, confirming the need for future model tuning.

A comparison with existing tools like SpeakEasy, AphasiaWeb, and SpeechBuddy shows that this platform offers deeper integration of AI, better gamification, and a future-ready therapist dashboard, making it more comprehensive and adaptable.

A prototype was developed and tested on a small group of speech-impaired individuals. Metrics such as pronunciation accuracy and speech fluency improvements were recorded. Initial results indicate increased patient motivation and measurable speech progress.

Testing Process:

Evaluation Strategy:

Word Error Rate (WER): Used to assess speech-to-text model accuracy.

Latency: Measures the system's response time in generating feedback.

Speech synthesis clarity: Evaluated through participant ratings on voice naturalness and intelligibility.

Participants: 10 speech-impaired individuals (ages 8–40).

Duration: 8 weeks of therapy, 3 sessions per week.

Metrics Tracked:

Pronunciation accuracy.

Speech fluency (words per minute).
Patient motivation (self-reported surveys).

Table 3. Pronunciation Accuracy Over Time

Weeks	Pronunciation Accuracy (%)
1	50
2	60
3	70
4	80
5	85
6	87
7	90
8	92

Key Observations:
The accuracy started at 50% in Week 1.
There was steady improvement each week.
By Week 8, accuracy reached 92%, indicating that the therapy had a significant impact on pronunciation skills.
The biggest improvement occurred between Week 3 and Week 5, where accuracy jumped from 70% to 85%.

Table 4. Speech Fluency Improvement

Participants	Pre-Test WPM	Post-Test WPM
1	40	65
2	42	67
3	39	64
4	41	66
5	43	68
6	38	63
7	44	69
8	40	65
9	42	67
10	39	64

Key Observations:
The Pre-Test WPM (before therapy) ranged between 38-44 WPM, with an average of 40 WPM.
The Post-Test WPM (after therapy) showed a clear improvement, ranging between 63-69 WPM, with an average of 65 WPM.
Every participant showed an increase of at least 25 WPM, demonstrating improved fluency after therapy.

Table 5. Patient Motivation

Category	Percentage
Very Satisfied	70%
Satisfied	20%
Neutral	10%

Key Observations:
70% of participants felt "Very Satisfied" with the therapy.
20% were "Satisfied", meaning they found it helpful but perhaps not perfect.
Only 10% were "Neutral", indicating that very few participants felt indifferent.
0% dissatisfaction, meaning no one found the therapy ineffective.

Conclusion and Future Work:

The proposed AI-powered speech therapy tool demonstrates significant potential [8] in enhancing the accessibility, affordability, and efficiency of speech therapy for individuals with speech impairments. By leveraging advanced generative AI techniques, the system provides personalized therapy exercises, real-time feedback, and adaptive learning, which can be particularly beneficial for those who lack access to traditional in-person therapy due to geographical, financial, or logistical constraints. The integration of artificial intelligence in speech therapy not only allows for continuous patient engagement but also fosters self-paced learning, ultimately improving speech fluency and confidence in communication.

Future work will focus on several key areas to further optimize and validate the effectiveness of the system. One primary objective is to expand the dataset by incorporating a more diverse range of speech samples, including different languages, accents, and severity levels of speech impairments. This will enable the AI model to generalize better across various patient demographics, ensuring inclusivity and improved accuracy in recognizing and responding to different speech patterns.

Another critical aspect of future development is refining the real-time feedback mechanism to enhance its precision and responsiveness. By integrating more sophisticated deep learning models and fine-tuning speech recognition algorithms, the system can provide more accurate and context-aware corrections, making therapy sessions more interactive and effective. Additionally, incorporating user-specific progress tracking and intelligent recommendations will further personalize the therapy experience, catering to everyone's unique challenges and improvement areas.

Moreover, to assess the clinical validity and real-world impact of the AI-powered therapy tool, large-scale clinical trials will be conducted in collaboration with healthcare institutions, speech-language pathologists, and rehabilitation centers. These trials will help gather empirical evidence on the system's effectiveness, user satisfaction, and long-term benefits, paving the way for potential integration into mainstream speech therapy practices. Future enhancements may also include multi-modal therapy support, integrating visual and haptic feedback, gamification elements to improve user engagement, and compatibility with assistive communication devices.

Table 6. Components and updated technology used data

Component	Updated Technology Used
Web Framework	React.js (Frontend), Node.js with Express (Backend)
Speech Recognition	Google Speech-to-Text API
Speech Synthesis	Amazon Polly (if used), or retaining TTS (Google TTS)
AI/NLP Engine	Amazon Lex, Hugging Face Transformers
Machine Learning Library	TensorFlow / PyTorch
Database	MongoDB
Authentication	JWT (JSON Web Tokens)
Hosting & Deployment	AWS, Firebase
Frontend Technologies	HTML, CSS, JavaScript
Component	Updated Technology Used
Web Framework	React.js (Frontend), Node.js with Express (Backend)
Speech Recognition	Google Speech-to-Text API
Speech Synthesis	gTTS (Google TTS)
AI/NLP Engine	Amazon Lex, Hugging Face Transformers
Machine Learning Library	TensorFlow / PyTorch
Database	MongoDB
Authentication	JWT (JSON Web Tokens)

Hosting & Deployment	AWS, Firebase
Frontend Technologies	HTML, CSS, JavaScript

Acknowledgments:

We are profoundly grateful to all those who helped us in this research project, especially the faculty and staff of the University of Engineering and Technology, Peshawar, Pakistan.

The manuscript has not been published or submitted to other journals previously.

Author's Contribution:

It is acknowledged that all authors have contributed significantly and that all authors agree with the content of the manuscript.

Conflict of Interest:

There exists no conflict of interest for publishing this manuscript in IJIST.

References:

[1] L. Mortensen, A. S. Meyer, and G. W. Humphreys, "Age-related effects on speech production: A review," *Lang. Cogn. Process.*, vol. 21, no. 1–3, pp. 238–290, Jan. 2006, doi: 10.1080/01690960444000278;JOURNAL:JOURNAL:PLCP20;WGROUP:STRING:PUBLICATION.

[2] D. Hirtz, D. J. Thurman, K. Gwinn-Hardy, M. Mohamed, A. R. Chaudhuri, and R. Zalutsky, "How common are the 'common' neurologic disorders?," *Neurology*, vol. 68, no. 5, pp. 326–337, Jan. 2007, doi: 10.1212/01.WNL.0000252807.38124.A3;PAGE:STRING:ARTICLE/CHAPTER.

[3] K. S. Regina Jokel, Luc De Nil, "Speech disfluencies in adults with neurogenic stuttering associated with stroke and traumatic brain injury," *J. Med. Speech. Lang. Pathol.*, vol. 15, no. 3, pp. 243–261, 2007, [Online]. Available: https://www.researchgate.net/publication/279587929_Speech_disfluencies_in_adults_with_neurogenic_stuttering_associated_with_stroke_and_traumatic_brain_injury#:~:text=The six features of neurogenic,positions%3B 4) the speaker does

[4] D. Ward, "Stuttering and cluttering: Frameworks for understanding and treatment," *Psychol. Press*, 2006, [Online]. Available: https://psycnet.apa.org/record/2006-20272-000

[5] P. van L. Aravind Kumar Namasivayam, Deirdre Coleman, Aisling O'Dwyer, "Speech Sound Disorders in Children: An Articulatory Phonology Perspective," *Front. Psychol.*, vol. 10, 2019, doi: https://doi.org/10.3389/fpsyg.2019.02998.

[6] S. T. Benjamas Prathanee, Panida Thanawirattananit, "Speech, language, voice, resonance and hearing disorders in patients with cleft lip and palate," *J Med Assoc Thai*, vol. 96, 2013, [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24386745/

[7] J. R. Duffy, "Chapter 33 - Functional speech disorders: clinical manifestations, diagnosis, and management," *Handb. Clin. Neurol.*, vol. 139, pp. 379–388, 2016, doi: https://doi.org/10.1016/B978-0-12-801772-2.00033-3.

[8] F. Brosseau-Lapré and S. Rvachew, "Introduction to speech sound disorders," *Plur. Publ.*, p. 295, 2020.

[9] R. S. Apurba Das, Nawed Alam Bhattacharjee, Akash Pal, Dipti, "Speech and Language Disorder: Assessment and Intervention Approaches," *Int. J. Soc. Impact*, vol. 9, no. 1, 2024, doi: 10.25215/2455/0901013.

[10] A. L. Holland, "When is Aphasia Aphasia? The Problem of Closed Head Injury," *Aphasiology Arch.*, 1982, [Online]. Available: https://aphasiology.pitt.edu/746/