

SpeakBoost: Enhancing Verbal Communication with AI

Aditya Dubey¹, Aditi Ambasta², Jenish Soni³, Shreya Singh⁴, Pranav Pawar⁵ and Sudhir Dhekane⁶

¹⁻⁶Artificial Intelligence and Data Science, Dwarkadas J. Sanghvi, College of Engineering, Mumbai, 400056, Maharashtra, India

Email: {aditya.work132, aditiambasta2004, jenishsoni62@gmail, shreyaraj1604, ppawar2612}@gmail.com, sudhir.dhekane@djsce.ac.in

Abstract— Effective communication is essential for success in both personal and professional settings. However, many individuals face challenges such as unclear articulation, nervousness, and difficulty maintaining audience engagement. Addressing these issues requires consistent practice and constructive feedback. This paper explores the integration of artificial intelligence (AI) to enhance communication skills through personalized, data-driven coaching. Our AI-powered platform, SpeakBoost, allows users to simulate real-world speaking scenarios, assessing performance based on key metrics such as speech clarity, pace, tone, filler usage, grammar errors. It generates comprehensive reports that highlight strengths, pinpoint areas for improvement, and provide actionable insights. With progress tracking over time, users can monitor their development, build confidence, and refine their communication style. By combining immersive practice with real-time feedback, this solution bridges the gap between theory and real-world application, empowering individuals to communicate more effectively in diverse contexts

Keywords— Artificial Intelligence, Natural language Processing, Communication, Human Computer Interaction.

I. INTRODUCTION



Fig. 1 Top 10 most In-Demand Soft Skills

Communication is more than just exchanging words. It involves conveying messages clearly, which leads to understanding and meaningful actions in both personal relationships and professional settings. This is essential for achieving success. Many people struggle with communication, but these skills can be developed and improved to increase both confidence and productivity. In today's knowledge-driven economy, the need for strong communication skills has grown. As shown in Fig. 1, a 2015 LinkedIn survey of 291 hiring managers in the U.S. found that 59% have trouble finding candidates with soft skills, including communication. Additionally, 58% think that a lack of these skills directly limits their company's productivity.

Niagara Institute, 2023 [2] states that effective communication boosts output and relationships at work. However, poor communication can lead to higher stress, low morale, missed performance goals, and decreased job satisfaction. These numbers show just how important effective communication is for your career:

- In 2023, leadership, communication, and customer service skills will be the top three most sought-after competencies, according to LinkedIn [3].
- Nearly 70% of remote workers experience burnout due to digital communication. In contrast, on-site and hybrid workers face less burnout from digital communication, at 56% and 49%, respectively, as reported by Forbes [4].
- People say that poor communication at work contributes to increased stress (52%), project delays or unfinished tasks (44%), reduced sales (18%), lower morale (31%), and missed performance targets (25%), according to Lucidchart [5].
- Knowledge workers indicate that inadequate communication makes them consider looking for a new job (22%), decreases their confidence (30%), lowers their job satisfaction (34%), and increases their stress (50%), as shown in Fig. 2. Effective communication, on the other hand, improves productivity, working relationships, and job satisfaction, as mentioned by Grammarly [6].



Fig. 2 Consequence of Poor Communication

In addition to workplace success, communication is fundamental for interpersonal relationships. Better relationships, deeper relationships, and resolving conflict are all achieved through empathetic and open communication. It has been found that individuals able to communicate their emotions have healthier social relationships and better mental health (Kapable, n.d. [7]).

This challenge is intensified by factors such as burnout due to online communication, cultural differences, and insufficient practice opportunities (Grammarly 2024 [6]; Forbes, 2023 [4]). New strategies that provide personalized feedback and adaptive learning opportunities are necessary to bridge this gap and experiences with adaptive learning.

A. Product: SpeakBoost

We present an AI-powered communication enhancement tool that provides immediate, tailored feedback in order to address the difficulties people encounter when learning professional and public communication. By emphasizing crucial components like clarity, tone, structure, pacing, and overall delivery, this creative approach fills the gaps left by conventional communication training techniques and aids users in improving their speaking abilities.

The application creates a thorough performance report that outlines the user's advantages and pinpoints particular areas in need of development. The platform lets users track their progress over time and keep an eye on improvements in important communication metrics by storing their performance data.

The tool uses state-of-the-art AI technologies to deliver personalized, accessible, and adaptive real-time feedback. In the end, this AI-powered solution changes the way people think about communication.

II. LITERATURE REVIEW

In their paper "A Voice-Input Voice-Output Communication Aid for People with Severe Speech Impairment," M. S. Hawley, S. P. Cunningham, et al. suggest a multimodal system that uses text, audio, and video inputs to evaluate communication skills. The results are presented as a downloadable report with graphical feedback and insights.

Its drawbacks include the fact that it only supports English, disregards grammatical correctness and vocabulary depth, and fails to take speech content into account. Feedback is static and does not provide tutorial-based or interactive guidance for incremental improvement. There are also no tracking features or customized profiles. [8]

In "Real-time Speaking Rate Monitoring System," M. A. Pandharipande, S. K. Koppurapu, et al. describe an AI-based communication skill development assistant for schoolchildren that processes voice and text using ANNs and TensorFlow, producing a performance report that compares inputs using data that has been trained. Its drawbacks include limited accuracy when processing real-time voice inputs, limited support for English, and no evaluation of grammar or vocabulary. These issues highlight issues with automation and performance reliability. [9]

In "Acoustic Analysis for Automatic Speech Recognition," D. O'Shaughnessy et al. present a multimodal automated oral English assessment model with a focus on parameter optimization through the use of CBOW and Hierarchical Softmax techniques. The need for better, larger datasets is highlighted by limitations such as the disparity in scale and quality between experimental and real-world data. For improved prediction efficacy and accuracy, future studies should compare algorithms and adjust parameters. [10]

In "Automatic Assessment Method of Oral English Based on Multimodality," X. Chen presents a framework that combines mouse movements, keystrokes, and facial expressions to gauge student engagement. The framework achieves 95.23% accuracy in identifying engagement level. It does not, however, provide real-time analysis, individualized reports, or tutorials to help students increase their level of engagement. [11]. In their paper "Semi-Automated Collaborative Online Training Module for Improving Communication Skills," R. Zhao, V. Li, H. Barbosa, G. Ghoshal, and M. E. Hoque introduce ROC Speak, a module that uses repeated practice over ten days to enhance public speaking abilities. Speaking abilities as evaluated by raters and peers were enhanced by feedback mechanisms. Limitation include low participant engagement and feedback diversity, as well as the lack of real-time feedback (users upload recorded videos). [12]

III. RELATED WORK

Needs for communication training are met by a variety of platforms. In comparison to SpeakBoost, this section examines the main current solutions and their drawbacks.

A. Applications of Speech Analysis

VR settings and AI feedback for public speaking are combined in Virtual Speech, which analyzes eye contact, pace, and filler words [14]. Accessibility is limited, though, because VR gear is needed.

A smartphone app called Orai offers automated speech coaching on pace, clarity, and filler words [15]. Unlike SpeakBoost, it does not have scenario-specific training or real-time dialogue.

B. AI Discussion Partners

Speeko offers AI speech coaching that emphasizes structured practice and delivery metrics [16], but it does not offer interactive AI conversation partners for real-time situations. Yoodli provides AI feedback on body language, pacing, and filler words in presentations [17]. It offers asynchronous feedback in the absence of scenario-based, real-time conversational AI capabilities.

C. Platforms for Enterprise Communication

Quantified Communications primarily targets corporate programs without interactive learning for individuals, but it also offers enterprise-level analytics measuring clarity, confidence, and engagement [18].

In order to enhance business communication, Gong.io examines sales conversations [19], emphasizing retrospective analysis over proactive training in a variety of contexts.

D. Research Deficit

Current solutions primarily fall into:

- Technical feedback devoid of context is provided by metric-focused analysers.
- Scenario simulators offer little room for customization.

- Important restrictions consist of absence of dynamic interaction, analysis at the surface level, and feedback that is generic and not very personalized.

E. The Innovative Contributions of SpeakBoost

SpeakBoost fills these voids by providing:

- Realistic, context-aware practice with interactive AI conversation partners.
- Multifaceted analysis that takes into account grammar, tone, clarity, confidence, and engagement.
- Web-based design that is accessible without the need for specialized hardware.

By combining these features, SpeakBoost represents a significant advancement in communication training technology, bridging the gap between generic speech analysis tools and personalized communication coaching.

IV. METHODOLOGY

A. User Research and Market Segmentation

User research and market segmentation were done to find speech-related issues and confirm the need for AI-driven coaching tools in order to make sure SpeakBoost tackles real-world communication challenges.

Target Audience

SpeakBoost is intended for both professionals and individuals who want to communicate more effectively verbally. Important sections consist of:

- HR & Talent Development Managers: Business executives who want to improve the leadership and communication abilities of their staff.
- Professional Workers: Business executives and early-career professionals who need to communicate effectively and confidently.
- Non-native speakers and immigrants: People looking to increase their confidence, pronunciation, and fluency in English.
- Podcast and Content Producers: Producers who need to improve their delivery, engagement, and speech clarity

User Interviews

Professionals from a variety of industries were interviewed to gain insight into the value of communication skills and the potential contribution of AI tools to their improvement.

Needs for Communication:

- Public speaking coaches, trainers, and speakers: In order to engage audiences, these professionals need to possess executive presence and sophisticated presentation skills.
- Educational institutions: Effective teaching, learning, and academic presentations require clear, coherent communication between instructors and students.
- Difficulties Met:
- HR managers: Find communication gaps that impact leadership development and promotions.
- Professional workers: Have trouble clearly expressing complicated concepts in presentations and group settings.
- Immigrants and non-native speakers: Face language hurdles and cultural quirks that prevent effective communication.
- Podcast and content creators: Require voice modulation, filler reduction, and audience engagement strategies.
- Current training methods and deficits: Organizations frequently employ buddy systems, workshops, and outside trainers. Although useful for fundamental abilities, these don't provide the individualized, real-time feedback required for positions requiring sophisticated speaking abilities or cultural sensitivity.

AI Tool Potential:

All user groups expressed clear interest in AI-driven communication tools:

- Instantaneous feedback on tone, filler words, speech rate, and grammar
- Support for intercultural communication, especially for international teams and non-native speakers
- Enhanced delivery, clarity, and pacing feedback for content creators

Considerations for Adoption:

Important elements influencing adoption include:

- Features tailored to different professional roles (e.g., podcast creators, HR managers)
- Cost-effectiveness and accessibility for learners and educational institutions

- Scalability for corporate training and large organizations

Future Development Steps:

Based on the findings, the following actions are recommended:

- Create customized modules (e.g., public speaking) for each target audience
- Include cultural communication guidelines and simulations
- Ensure cultural sensitivity, especially for global users and immigrants
- Develop feedback systems compatible with various media types, including podcasts, interviews, and presentations

These initiatives will guide SpeakBoost's development to remain practical, adaptable, and tailored to the unique needs of each professional group.

Market Segmentation

Effective adoption of SpeakBoost requires an understanding of its diverse audience:

- With the rapid adoption of AI, the language learning market is expected to reach \$21 billion by 2025 [19].
- The social media market is projected to hit \$122.6 billion by 2025, indicating rising demand for speech coaching among content creators [20].
- The corporate training market is projected to reach \$370 billion by 2025, emphasizing the need for leadership and communication skills [21].
- The online education market is expected to reach \$350 billion by 2025, highlighting AI's growing role in digital learning [22].

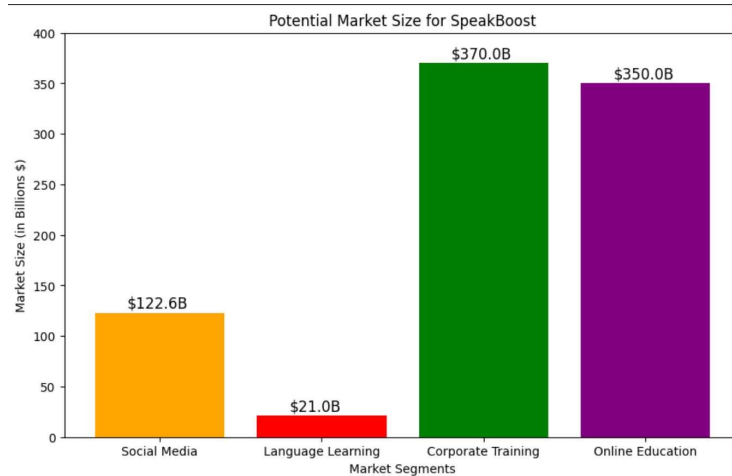


Fig. 3 Potential Market Size for SpeakBoost

B. Technology Stack and System Development Methodology

Through accurate, adaptive, and real-time feedback based on strong linguistic analysis and AI-driven insights, Speak Boost aims to improve users' spoken communication abilities. The system is designed to work as an end-to-end web-based platform that records and transcribes user speech while also analysing and contextualizing it to produce helpful criticism meant to enhance speech effectiveness in general as well as fluency, pacing, and grammar. Modern web application frameworks, cutting-edge artificial intelligence (AI) services, and dependable data persistence mechanisms must all be seamlessly integrated into a unified technology stack that is optimized for scalability and interactivity.

The SpeakBoost pipeline flowchart is displayed in Fig. 4. The platform is built around the Flask framework [23], which was chosen for its lightweight implementation, simple design, and robust support for RESTful API development. These features make it perfect for creating and administering asynchronous, modular back-end services. Flask's secure routing and session control features are used to implement user authentication, session tracking, and scenario management. MySQL [24], a powerful relational database system, manages persistent data storage. It has high consistency and transactional reliability in storing user credentials, session logs, speech analysis reports, and user feedback data.

Speak Boost uses OpenAI's Whisper API [25] for speech input processing, a transformer-based automatic speech recognition (ASR) model renowned for its multilingual robustness and high transcription accuracy. noise level or

a variety of accents. OpenAI's GPT-4 [26], which powers tasks like grammar correction, sentiment and tone analysis, and the creation of personalized performance feedback through structured natural language prompts, is used to evaluate user speech after it has been transcribed. By producing interactive dialogue flows and context-aware questions, the model also plays a significant part in scenario simulation. The system incorporates Google's Text-to-Speech (TTS) API [27] to deliver real-time auditory feedback, allowing for the delivery of synthetic speech responses in an expressive and clear voice. Combining text and speech in this multimodal feedback loop improves accessibility and user engagement, particularly for auditory learners.

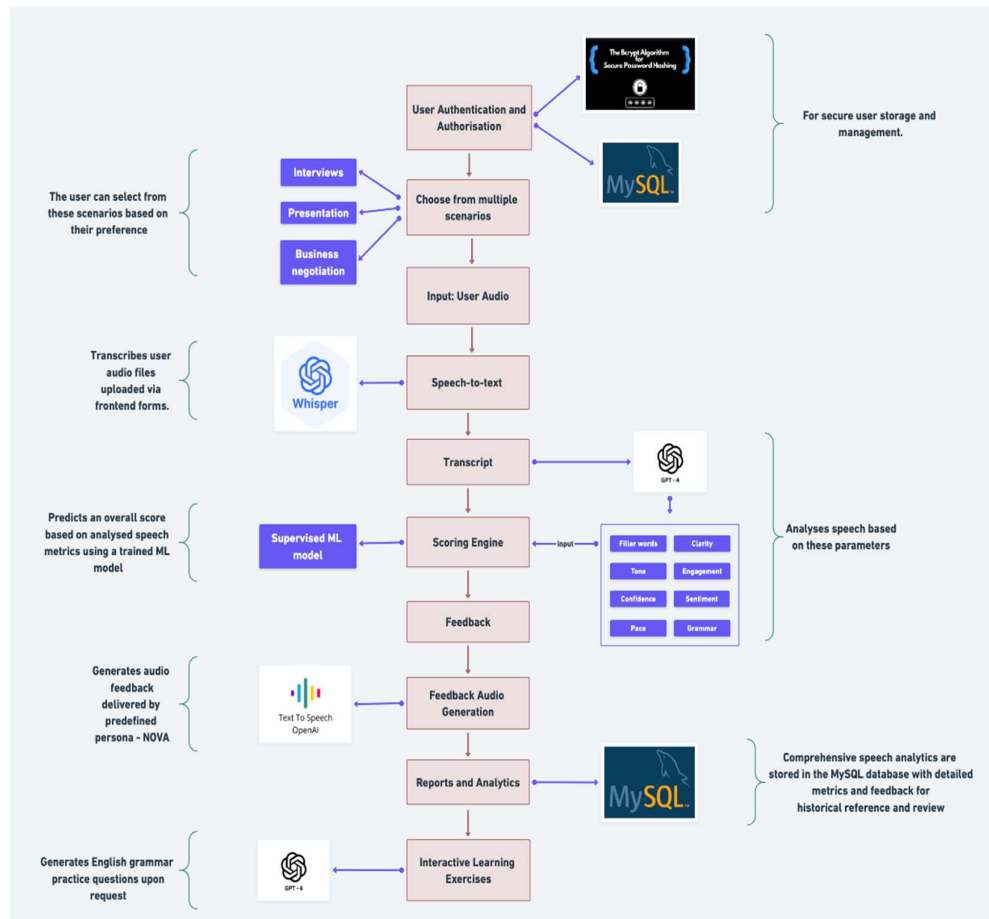


Fig. 4 Pipeline Flowchart of SpeakBoost

In parallel, low-level audio features are analysed using the Librosa audio processing library [28], which makes it possible to calculate speaking rate (words per minute) and speech duration—two important fluency indicators. These parts work together in a closely coordinated pipeline that starts with browser-based audio recording, moves on to Whisper transcription, GPT-4-driven speech evaluation, and ends with TTS-delivered auditory feedback. MySQL makes sure that all interactions, session data, and feedback iterations are kept on file so that users can track their progress and get customized practice resources derived from past performance. A high-level architectural overview of the system is shown in Fig. 4, which also highlights the coordination between each platform subsystem and shows the functional flow from speech input to final feedback generation.

Core Functionalities and Implementation

- **Authorization & Authentication of Users:** The platform uses Werkzeug security tools and Flask session management to implement role-based access control and secure user login. For added security, hashed passwords are stored securely in the MySQL database [24]. A session cookie is sent by Flask after successful authentication, tracking and maintaining the user's logged-in status throughout requests. The

application has strong foundational security through standard Flask approaches such as decorators for login enforcement and blueprints for modular route management.

- **Scenario-Based Interactive Learning:** SpeakBoost offers simulated situations such as business presentations, job interviews, and casual conversations to support scenario-driven practice sessions. Prompts, context descriptions, and dynamic dialogue scripts are dynamically served through the web interface and stored in the MySQL database. Users respond to customized prompts during interactive sessions by speaking into their devices. To create realistic, scenario-relevant follow-up questions or to mimic an interactive interlocutor, the system leverages GPT-4 [26].

Users are provided with an immersive and customized learning environment through this iterative process.

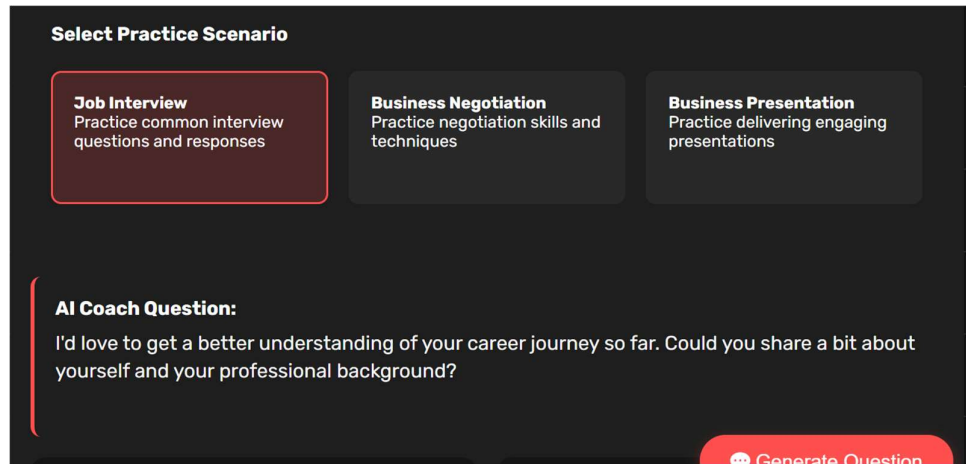


Fig. 5 Practice Scenarios

- **Whisper API Speech-to-Text Conversion:** OpenAI's Whisper API is used to convert user-provided audio inputs into text via browser-based interfaces (using the Web Audio API) [25]. Whisper uses a strong Transformer-based encoder-decoder architecture that is especially tuned for efficient noise management and high accuracy in recognizing a variety of accents.

RESTful API calls are used to send audio data to Whisper, which returns transcriptions in JSON format. The Speak Boost system uses this textual transcription as the foundation for further linguistic and acoustic analyses.

```
# Transcribe with Whisper
with open(temp_audio.name, 'rb') as audio:
    transcript_response = client.audio.transcriptions.create(
        model="whisper-1",
        file=audio
    )
transcript = transcript_response.text
```

Fig. 6 Speech-to-Text Conversion

Calculating Words Per Minute (WPM)

Words per minute (WPM) is a quantitative indicator of speech fluency. The WPM metric is calculated as follows, where N is the transcript's total word count and T is the audio duration in seconds:

$$\text{WPM} = \frac{N}{T/60} \quad (1)$$

Speak Boost makes use of the Librosa audio processing library to measure duration accurately [28]. In order to enable longitudinal fluency tracking across user sessions, the generated WPM metric is logged.

- Grammatical accuracy and the minimal use of filler words are essential for clear spoken communication. To detect grammatical deviations in user transcripts and provide corrections or improved phrasing suggestions, GPT-4 [26] is utilized. In addition, keyword-based transcript scanning procedures, which leverage the explicit Whisper transcription of verbal hesitations, are employed to identify filler words such as “um,” “uh,” “like,” and “you know.”
- Customized Feedback Production Speak Boost creates individualized user feedback through GPT-4 by synthesizing analysed metrics (fluency, sentiment, grammar, and filler words) [26]. GPT-4's advanced text-generation features enhance user motivation and clarity toward areas for improvement by customizing feedback to user performance, contextual scenario relevance, and communicative style
- Customized Feedback Production Speak Boost creates individualized user feedback through GPT-4 by synthesizing analyzed metrics (fluency, sentiment, grammar, and filler words) [26]. GPT-4's advanced text-generation features enhance user motivation and clarity toward areas for improvement by customizing feedback to user performance, contextual scenario relevance, and communicative style.
- Text-to-Speech (TTS) Audio Feedback In order to improve user involvement and the Google Text-to-Speech (TTS) API is used to convert textual feedback to audio for accessibility [27]. The GPT-4-generated post-analysis textual feedback is asynchronously transformed into synthetic speech that sounds natural, this multimodal delivery preserves an immersive learning environment while assisting users in intuitively understanding feedback nuances.
- Analytics and Reporting in the MySQL database, the platform methodically logs comprehensive session metrics, such as speech duration, computed WPM, filler word frequency, sentiment analysis, and GPT-generated feedback summaries [24]. Through an interactive frontend dashboard that shows progress over several sessions, users can examine their performance analytics. Client-side JavaScript chart libraries facilitate visualization for an engaging user interface
- Grammar Tasks and Creating Questions Speak Boost uses GPT-4 to dynamically generate focused grammar exercises and thoughtful follow-up questions to improve linguistic accuracy [26]. The system logs generated exercises and answers, allowing for continuous skill reinforcement and repeated practice outside of scheduled

Iterative Feedback Improvement Driven by Supervised Learning Speak Boost uses an iterative refinement mechanism inspired by supervised learning to methodically optimize feedback produced by GPT-4, guaranteeing exactness, succinctness, and comprehensiveness. This iterative evaluation-and-adjustment method is very similar to supervised learning paradigms in that outputs are assessed against predetermined criteria, resulting in prompt improvement and correction.

Metrics for Assessing the Quality of Feedback

Three different metrics are used to objectively evaluate the quality of feedback in Speak Boost:

- Relevance (R): Assures contextual accuracy by gauging how closely the feedback matches the user's actual spoken transcript.
- Conciseness (C): Assesses how concisely the feedback is delivered, promoting a length that successfully conveys the required edits without being overly wordy.
- Completeness (M): Verifies that the feedback fully covers all crucial areas determined by user performance, including overall delivery, speech fluency, and grammatical accuracy.

Every metric is calculated algorithmically using the original feed produced by GPT-4, returned and normalized using a 0–1 scale. These three metrics are averaged to determine the final combined accuracy score (accuracy):

$$\text{accuracy} = \frac{R+M+C}{3} \quad (2)$$

Feedback quality can be succinctly and quantitatively measured with this single scalar score.

The Process of Iterative Refinement

These specific steps are systematically followed by the iterative feedback refinement process inspired by supervised learning:

- First Prediction (Generation GPT-4): GPT-4 creates the first feedback text using the scenario context and the user's transcribed speech.
- Performance Evaluation (Accuracy Scoring): To calculate Saccuracy, the system compares this feedback to the predetermined accuracy metrics (Completeness, Conciseness, and Relevance).
- Choice (Threshold Evaluation): If the accuracy meets or exceeds the threshold, the user receives the feedback directly as it meets quality standards; otherwise, if accuracy is below the threshold, the feedback is

considered insufficient and explicit refinement instructions are generated to guide GPT-4 on the necessary improvements.

- Refined Prediction (Feedback Regeneration): GPT-4 effectively supervises the explicit accuracy criteria while generating the feedback in accordance with these explicit, guided instructions.
- Final Evaluation: Iteratively repeating steps 2-4 (usually 1-2 iterations) quickly converge to feedback that meets or surpasses the threshold τ .
- The refinement loop typically demonstrates efficiency and dependability by achieving acceptable feedback accuracy in a small number of iterations. The iterative feedback refinement mechanism is graphically depicted in Fig. 7.

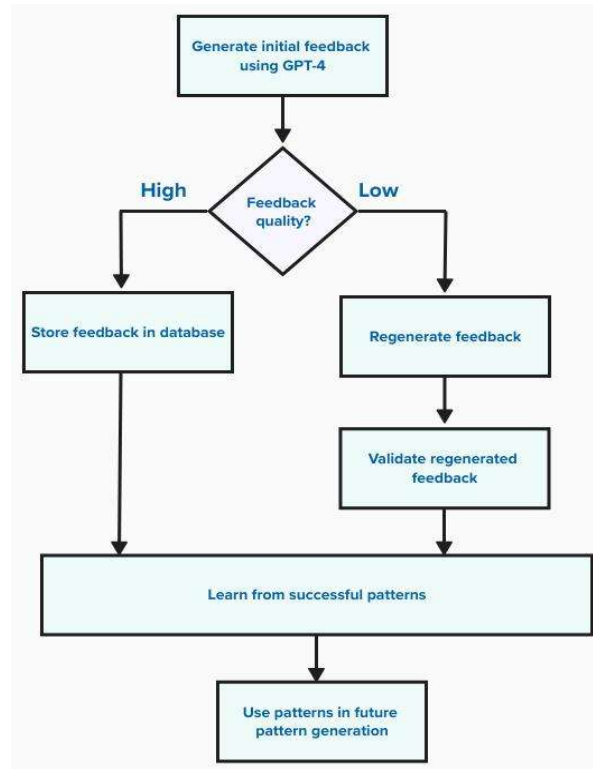


Fig. 7 Supervised Learning-Inspired Iterative Feedback Refinement Loop

Similar to supervised learning, this evaluation-and-correction method compares results to predetermined goals for instantaneous improvement. The use of accuracy metrics allows for a supervised-like refinement that improves feedback quality even though GPT-4 isn't retrained.

V. RESULTS

The architecture, operation, and possible societal uses of the SpeakBoost system are presented in this section.

A. Overview of the System

With real-time, tailored feedback, SpeakBoost is an AI-powered platform that helps users improve verbal communication. The system incorporates GPT-4 for linguistic analysis and OpenAI's Whisper model for speech recognition. The user starts the process by using the SpeakBoost web interface to record their speech. Whisper transcription of the audio is followed by a GPT-4-based assessment for grammar feedback, speech pacing analysis, and filler word detection. The system then displays a thorough feedback report that includes the following information:

- Total word count Filler words were identified, and suggested pacing changes were made.
- Suggestions for improving grammar Users can instantly examine their speech patterns and work to improve their verbal fluency thanks to this feedback.

B. Key Features

- Real-time AI feedback: Speech from users is instantly processed and analysed.
- User-Friendly Interface: Flask-powered platform that is easy to use and accessible.
- Focused Insights: Pay attention to grammar, pacing, and filler words.
- Scenario Based Learning

C. Important System Elements

- Language Learning: Students receive immediate support to enhance their verbal abilities.
- Public Speaking Training: Helps users minimize filler usage and improve speech fluency

D. Expected Advantages

- For those without access to official training materials, offer easily accessible communication coaching.
- Provide diverse populations with scalable, AI-powered speech improvement assistance.

VI. CONCLUSION

Many individuals face challenges in speaking clearly, confidently, and fluently, even though effective communication is essential in numerous fields. This paper presented SpeakBoost, an AI-powered speech coaching tool that offers instantaneous feedback on spoken communication. By leveraging advancements in speech recognition, natural language processing (NLP), and deep learning, SpeakBoost analyses speech patterns to identify filler words, measure speech speed, assess sentiment, and provide grammatical corrections.

The platform integrates modern machine learning models implemented in Python, PyTorch, and transformers, alongside GPT-4 and Flask for deployment, with an intuitive web interface developed using HTML, CSS, and JavaScript. Experimental results demonstrate high accuracy in sentiment analysis, speech pacing, and grammatical corrections, positioning SpeakBoost as a reliable tool for enhancing speaking abilities.

Through market segmentation and research, key target groups have been identified, including non-native speakers, corporate professionals, educational institutions, and podcast creators. Future work will focus on improving model accuracy, expanding multilingual support, and incorporating personalized learning pathways. By enabling users to enhance their speech with confidence, clarity, and precision, SpeakBoost represents a significant step toward bridging the gap between artificial intelligence and human communication.

FUTURE SCOPE

There is considerable potential for further development and enhancement of this AI-powered communication assistant. Future improvements may include integrating video analysis to assess eye contact, facial expressions, and body language for feedback on nonverbal communication; introducing an interactive practice mode that allows users to rehearse optimized versions of their transcriptions with a focus on grammar, tone, and tempo; enabling real-time platform integration to provide immediate feedback during live sessions on platforms such as Microsoft Teams, Zoom, and Google Meet; and expanding cross-language support to enhance multilingual capabilities, ensure translation accuracy, and improve accessibility for non-native speakers.

REFERENCES

- [1] LinkedIn, “Most In-Demand Soft Skills,” 2015. [Online]. Available: <https://www.linkedin.com/business/talent/blog/talent-strategy/most-indemand-soft-skills>. [Accessed: Jul. 6, 2025].
- [2] Niagara Institute, “The Importance of Communication Skills,” 2023. [Online]. Available: <https://www.niagarainstitute.com/blog/importance-of-communication-skills>. [Accessed: Jul. 6, 2025].
- [3] LinkedIn, “Most In-Demand Skills.” [Online]. Available: <https://www.linkedin.com/business/learning/blog/top-skills-and-courses/most-in-demand-skills>. [Accessed: Jul. 6, 2025].
- [4] Forbes, “Digital Communication in the Workplace.” [Online]. Available: <https://www.forbes.com/advisor/business/digital-communication-workplace/>. [Accessed: Jul. 6, 2025].
- [5] Lucidchart, “Miscommunications at Work Impact the Bottom Line.” [Online]. Available: <https://www.prnewswire.com/news-releases/miscommunications-at-work-impact-the-bottom-line-study-finds-300620637.html>. [Accessed: Jul. 6, 2025].
- [6] Grammarly, “State of Business Communication Report,” 2024. [Online]. Available: <https://go.grammarly.com/2024-state-of-business-communication-report>. [Accessed: Jul. 6, 2025].
- [7] Kapable, “The Importance of Communication Skills.” [Online]. Available: <https://kapable.club/blog/communication-skills/what-is-the-importance-of-communication-skills/>. [Accessed: Jul. 6, 2025].

- [8] M. S. Hawley, S. P. Cunningham, et al., “A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 119–127, Jan. 2013.
- [9] M. A. Pandharipande, S. K. Kopparapu, et al., “Real-time speaking rate monitoring system,” in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (SPCC)*, 2011, pp. 1–4.
- [10] D. O’Shaughnessy, et al., “Acoustic Analysis for Automatic Speech Recognition,” *Proc. IEEE*, vol. 101, no. 5, pp. 1038–1053, Apr. 2013.
- [11] X. Chen, “Automatic Assessment Method of Oral English Based on Multimodality,” *Scientific Programming*, vol. 2022, Article ID 8459273, pp. 1–11, Jan. 2022.
- [12] R. Zhao, V. Li, H. Barbosa, G. Ghoshal, and M. E. Hoque, “Semi-Automated Collaborative Online Training Module for Improving Communication Skills,” Jun. 30, 2017.
- [13] VirtualSpeech, “VR Public Speaking and Communication Skills,” *VirtualSpeech*, vol. 2023, pp. 1–5, Jan. 2023.
- [14] Orai, “AI Speech Coach for Public Speaking,” *Orai Technical Documentation*, vol. 2022, pp. 1–8, Mar. 2022.
- [15] Speeko, “AI Communication Coach,” *Int. J. Commun. Technol.*, vol. 14, no. 2, pp. 112–125, Apr. 2023.
- [16] Yoodli, “AI Speech Coach: Analysis and Applications,” *Commun. Technol. Rev.*, vol. 8, no. 4, pp. 67–82, Jul. 2023.
- [17] Quantified Communications, “Enterprise Communication Analytics,” *Bus. Commun. Q.*, vol. 45, no. 3, pp. 205–218, Sep. 2022.
- [18] Gong.io, “Revenue Intelligence Platform: Speech Analytics for Business,” *J. Bus. Intell.*, vol. 12, no. 2, pp. 78–94, Feb. 2023.
- [19] HolonIQ, “Language Learning Market to Reach \$21B by 2025,” 2025. [Online]. Available: <https://www.holoniq.com>. [Accessed: Jul. 6, 2025].
- [20] Statista, “Speech Recognition Market to Reach \$8.58B by 2025,” 2025. [Online]. Available: <https://www.statista.com>. [Accessed: Jul. 6, 2025].
- [21] Training Industry Report, “Corporate Training Market to Hit \$370B by 2025,” 2025. [Online]. Available: <https://www.trainingindustry.com>. [Accessed: Jul. 6, 2025].
- [22] Global Market Insights, “Online Education Market Growth,” 2025. [Online]. Available: <https://www.gminsights.com>. [Accessed: Jul. 6, 2025].
- [23] Pallets Projects, “Flask Documentation,” 2023. [Online]. Available: <https://flask.palletsprojects.com>. [Accessed: Jul. 6, 2025].
- [24] Oracle Corporation, “MySQL 8.0 Reference Manual,” 2023. [Online]. Available: <https://dev.mysql.com/doc/refman/8.0/en>. [Accessed: Jul. 6, 2025].
- [25] A. Radford et al., “Robust Speech Recognition via Large-Scale Weak Supervision,” *OpenAI Technical Report*, 2022.
- [26] OpenAI, “GPT-4 Technical Report,” *arXiv preprint arXiv:2303.08774*, 2023.
- [27] Google, “Cloud Text-to-Speech API Documentation,” 2023. [Online]. Available: <https://cloud.google.com/text-to-speech>. [Accessed: Jul. 6, 2025].
- [28] B. McFee, et al., “librosa: Audio and Music Signal Analysis in Python,” in *Proc. 14th Python in Science Conf.*, Austin, TX, 2015, pp. 18–25.