Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

A Machine Learning Framework for Automated Interview Feedback and Performance Evaluation

Savitha G
Dept. of CSE, RVITM Bengaluru, India
savithag.rvitm@rvei.edu.in

Abhay Rupesh
Dept. of CSE, RVITM Bengaluru, India
abhayrups@gmail.com

Aditi Rao H
Dept. of CSE, RVITM Bengaluru, India
aditinagesh03@gmail.com

Anoushka Saini
Dept. of CSE, RVITM Bengaluru, India
anoushkadevelops@gmail.com

Anna Sri Ranga Ganesh Anuraag
Dept. of CSE, RVITM Bengaluru, India *anuraagranga03@gmail.com*

***Abstract*—**The Smart Interview Coach is a cutting-edge plat- form designed to enhance interview preparation by providing personalized, actionable feedback on candidates' performance. Utilizing advanced machine learning, natural language process- ing, and audio analysis techniques, the system transcribes audio responses into text, evaluates vocal features such as pitch, speech rate, loudness, and pauses, and conducts sentiment analysis. It also detects filler words and assesses the quality of language used, including grammar, coherence, and vocabulary. At the conclusion of each session, the system generates a comprehensive feedback report, offering insights into strengths and areas for improvement. By offering detailed, real-time analysis, the Smart Interview Coach helps candidates build confidence, refine their communication skills, and improve their chances of success in interviews, making it a valuable tool for both job seekers and recruiters.

***Index Terms*—**AI-driven assessment, Filler word detection, HR interview evaluation, interview feedback, librosa, RoBERTa, sentiment analysis, Smart Interview Coach, speech analysis, transformer-based transcription

I. INTRODUCTION

In today's highly competitive job market, candidates face immense pressure to perform well in interviews, as these eval- uations are often the deciding factor in securing desired roles. Traditional methods of interview preparation, such as mock interviews and self-assessments, frequently lack the depth and personalization required to offer meaningful insights. These methods usually focus on the content of answers but fail to address key aspects of communication like speech fluency, emotional tone, or body language. As a result, candidates often remain unaware of their communication weaknesses, leaving them ill-prepared for the demands of high-stakes interviews.

To address this gap, the Smart Interview Coach has been developed, a cutting-edge system designed to evaluate and enhance a candidate's interview performance. This system leverages state-of-the-art technologies in Machine Learning (ML), Natural Language Processing (NLP), and audio analysis to provide personalized feedback based on the candidate's speech during an interview. By analyzing key vocal features such as pitch, loudness, speech rate, and pauses, as well as the emotional tone and linguistic quality of responses, the Smart Interview Coach offers actionable insights that go far beyond traditional feedback. The system operates through a seamless process: candidates record their responses to HR-related ques- tions, categorized by experience level, and the Smart Interview Coach transcribes the audio into text using a transformer-based model. The system then evaluates the transcribed speech for various vocal attributes using Librosa for audio analysis and applies sentiment analysis via a RoBERTa-based model. This allows it to detect not only the content of the response but also emotional nuances and any filler words such as "um" or "like," which can detract from the clarity and professionalism of the candidate's delivery.

Once the analysis is complete, the system generates a detailed feedback summary that highlights the candidate's strengths and identifies areas for improvement. This feed- back includes personalized, actionable suggestions to help candidates refine their communication skills, enhance their confidence, and perform better in future interviews. The goal is not just to improve interview performance but to empower candidates with the tools they need for long-term success in the professional world. By providing continuous, data- driven feedback throughout the interview process, the Smart Interview Coach introduces a revolutionary approach to inter- view preparation. It bridges the gap between traditional self- assessment and professional coaching, making it accessible to individuals across various fields. This system aims to support a wide range of users, from fresh graduates seeking their first role to experienced professionals preparing for senior positions.

Furthermore, the system's adaptability and scalability offer significant potential for future enhancement. As interview practices evolve, additional features like advanced emotion detection, support for multiple languages, and even video analysis can be integrated, ensuring the Smart Interview Coach remains relevant and valuable in the changing landscape of job interviews. Through its innovative use of technology, this system promises to not only enhance interview readiness but also provide a lasting impact on the way candidates prepare for one of the most critical aspects of career advancement.

II. RELATED WORKS

Accurate assessment of interview responses plays a pivotal role in improving candidate evaluation and feedback processes. With the integration of ML and NLP, modern systems have the capability to analyze communication skills, confidence levels, and content relevance more effectively than traditional methods. These technologies have facilitated advancements in sentiment analysis, emotion recognition, and automated interview systems, enabling precise evaluation of speech, emo- tions, and textual responses. This literature survey examines significant research contributions in these areas, exploring their methodologies, advantages, and challenges to provide insights into the evolving field of automated interview analysis.

Adaptive strategies in interviews, guided by multimodal social signals, have demonstrated potential in improving re- sponse quality and interaction personalization. By tailoring questions to participants' willingness to speak, these systems enhance information gathering, though challenges remain in achieving high accuracy and managing delays due to complex data extraction processes [1]. Similarly, the use of avatar-based feedback in training environments has been shown to reduce anxiety and improve communication skills, though generaliz- ability is limited by sample size and controlled conditions [2]. Integrating audio emotion detection with text sentiment analysis has proven effective for real-time sentiment evalu- ation, offering applications in disaster management and other time-sensitive scenarios. However, audio sentiment analysis often introduces noise due to high variance, and large-scale implementations face significant computational challenges [3]. Automated HR interview systems employing deep learning and NLP further enhance efficiency in candidate assessments, reducing hiring bias and improving candidate experience. Nev- ertheless, these systems require large datasets for training and careful ethical considerations to ensure fairness and privacy [4].

Chatbots have been utilized to support inclusive learning by providing accessible and confidential assistance, partic- ularly for disadvantaged students. These systems improve interactivity and accessibility but face limitations in handling sensitive scenarios and ethical challenges associated with machine learning capabilities [5]. Advances in NLP robustness highlight the importance of adversarial training, data augmen- tation, and multi-task learning in defending against attacks and ensuring reliable performance. Despite these advancements, research gaps remain in addressing multimodal robustness and practical application examples [6].

Interview bots leveraging NLP and ML have shown promise in competency assessment, offering cost-effective, biasreduced, and remotely usable solutions. However, lim- itations in data variability and training set size impact per- formance accuracy [7]. NLP methods for natural language understanding and generation enable applications in machine translation, sentiment analysis, and chatbots, improving in- formation extraction and human-computer interaction. Yet, challenges like context understanding, language ambiguity, and adaptability to new domains persist [8].

Audio-visual sentiment analysis has benefited from trans- fer learning techniques, where knowledge from textual data improves performance across modalities. This approach is well-suited for real-world applications but requires extended training times and is sensitive to dataset characteristics [9]. Systems integrating audio and visual cues for evaluating interview performance improve accuracy and efficiency but face computational challenges in real-time processing and limitations in emotional state recognition [10].

Automated personality recognition during video interviews, using deep learning frameworks like TensorFlow, provides scalable and efficient personality assessments. However, these models often suffer from dataset limitations, reducing their generalizability and applicability in diverse settings [11]. Sen- timent analysis methods utilizing machine learning effectively classify emotions and polarity in textual data, but challenges include small datasets, limited focus on complex classifica- tions, and lack of big data techniques [12].

III. DATASET

Dataset The dataset consists of a comprehensive collection of interview questions categorized according to the experience level of the candidate, namely basic, experienced, and both. The basic category includes questions that assess general communication skills, motivation, and personal traits, typically aimed at entry-level candidates. The experienced category focuses on evaluating candidates with more professional ex- perience, aiming to assess their problem-solving abilities, leadership skills, and industry knowledge. The questions clas- sified under the ”both” category are suitable for candidates of all experience levels and are designed to measure key competencies such as emotional intelligence, adaptability, and professionalism.

The dataset is in CSV format, containing 62 samples with a total size of 15KB, structured with 62 rows and 2 columns. This dataset plays a pivotal role in training machine learning models and automated systems designed to assess interview responses. These systems utilize the dataset to analyze various aspects of a candidate’s answer, including clarity, coherence, and emotional tone. The analysis can evaluate the candidate’s communication skills, professional demeanor, and emotional engagement, providing an objective, data-driven approach to candidate evaluation.

Furthermore, the dataset facilitates the creation of systems that can generate real-time feedback, improving the overall efficiency and effectiveness of the interview process by of- fering consistent, unbiased evaluations. Another significant application of this dataset lies in the development of sentiment analysis and emotion detection models. By examining the tone and content of the responses, machine learning algorithms can assess the emotional state of the candidate, whether they exhibit signs of stress, confidence, or engagement. This can enhance the interview experience by offering deeper insights into the candidate’s mental and emotional state during the conversation.

However, despite its usefulness, the dataset has limitations, such as its relatively narrow scope of questions and the lack of non-verbal data like body language or facial expressions, which could further enrich the analysis and provide a more holistic understanding of a candidate’s performance in an interview setting. Additionally, the dataset can be expanded to include diverse question types and topics to better simulate real-world interview scenarios across various industries. This expansion would enable more robust training for AI models.

IV. METHODOLOGY

The Smart Interview Coach follows a structured method- ology to evaluate and enhance a candidate’s interview perfor- mance shown in Fig 1. The system processes recorded speech, extracts key audio and text features, and employs advanced AI models to generate real-time feedback. The methodology consists of three key stages: Data Preprocessing, Feature Extraction, and Model Design, ensuring efficient, scalable, and accurate assessment.

A. Data Preprocessing

The data preprocessing stage prepares recorded speech and transcribed text for analysis by eliminating inconsistencies and ensuring data quality. The recorded audio undergoes noise reduction, volume normalization, and resampling to improve clarity and reduce environmental distortions. A transformer- based speech-to-text model is then used to transcribe re- sponses, preserving linguistic accuracy and context.

Following transcription, text preprocessing is performed to refine the extracted text. Stopwords, filler words, and redundant pauses are filtered out, while punctuation and formatting inconsistencies are corrected. These preprocessing steps ensure clean and structured input data, which is essential for effective feature extraction and response evaluation.

B. Feature Extraction

Once the data is preprocessed, feature extraction is per- formed to analyze both audio and textual characteristics of candidate responses. Audio features help evaluate vocal clarity, tone, and fluency, while text features focus on linguistic quality, sentiment, and relevance. By extracting these features, the system gains valuable insights into both how a candidate speaks and what they convey in their responses.

- Audio Feature Extraction

Using Librosa, the system extracts key vocal character- istics, including pitch, loudness, speech rate, and pauses. Pitch analysis helps assess tonal variation, which influ- ences confidence perception. Loudness measurements en- sure appropriate volume for speech clarity, while speech rate analysis determines if the candidate speaks too fast or too slow. Pause detection identifies hesitation patterns and fluency levels. These audio features provide valuable insights into the candidate's communication delivery and help generate feedback focused on speech clarity and effectiveness.

Text Feature Extraction

NLP techniques analyze transcribed responses to assess sentiment, coherence, and relevance. Sentiment analysis, powered by RoBERTa, determines whether responses exhibit confidence, neutrality, or hesitation, providing in- sights into engagement levels. Semantic similarity analy- sis, using Google Gemini Large Language Model (LLM), measures alignment between the candidate's response and an ideal answer. Additional text-based metrics, such as grammatical accuracy, structural coherence, and com- pleteness, further refine the evaluation.

By extracting both audio and text-based features, the system ensures a comprehensive assessment of a candidate's interview response. These extracted features provide data-driven insights into speaking patterns, response structure, and overall effec- tiveness, forming the foundation for personalized feedback generation.

C. Model Design

The Smart Interview Coach is built using a modular AI- driven architecture, integrating multiple models for speech- to-text transcription, audio analysis, NLP evaluation, and AI- generated feedback.

- Speech Processing and Analysis

The transformer-based speech-to-text model ensures ac- curate transcription, while the Librosa-based module ex- tracts key vocal parameters. These components work together to objectively evaluate communication delivery, measuring clarity, fluency, and tone.

- NLP and AI-Based Evaluation

Once the transcription is completed, RoBERTa performs sentiment analysis, while Google Gemini LLM mea- sures semantic similarity to assess response relevance. A Generative AI model synthesizes structured feedback, highlighting strengths and areas for improvement. This multi-layered analysis ensures that responses are assessed not only for delivery but also for content quality and engagement.

- Feedback Generation and System Optimization

The system compiles extracted insights into a structured feedback report that provides candidates with a detailed evaluation of their responses. The speech analysis com- ponent offers insights into clarity, fluency, and vocal delivery, helping users refine their speaking style and minimize hesitation. The text-based evaluation assesses the relevance, coherence, and sentiment of responses, ensuring that candidates' answers align with professional communication standards. Finally, the AI-generated feed- back module synthesizes these evaluations to deliver targeted improvement suggestions, enabling candidates to refine their speaking and response structure.

To ensure efficiency, the system is optimized for real-time feedback generation, allowing candidates to receive instant, actionable insights. The architecture is designed to scale efficiently, supporting multiple users simultaneously while

Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 6 Issue 1 (2026)

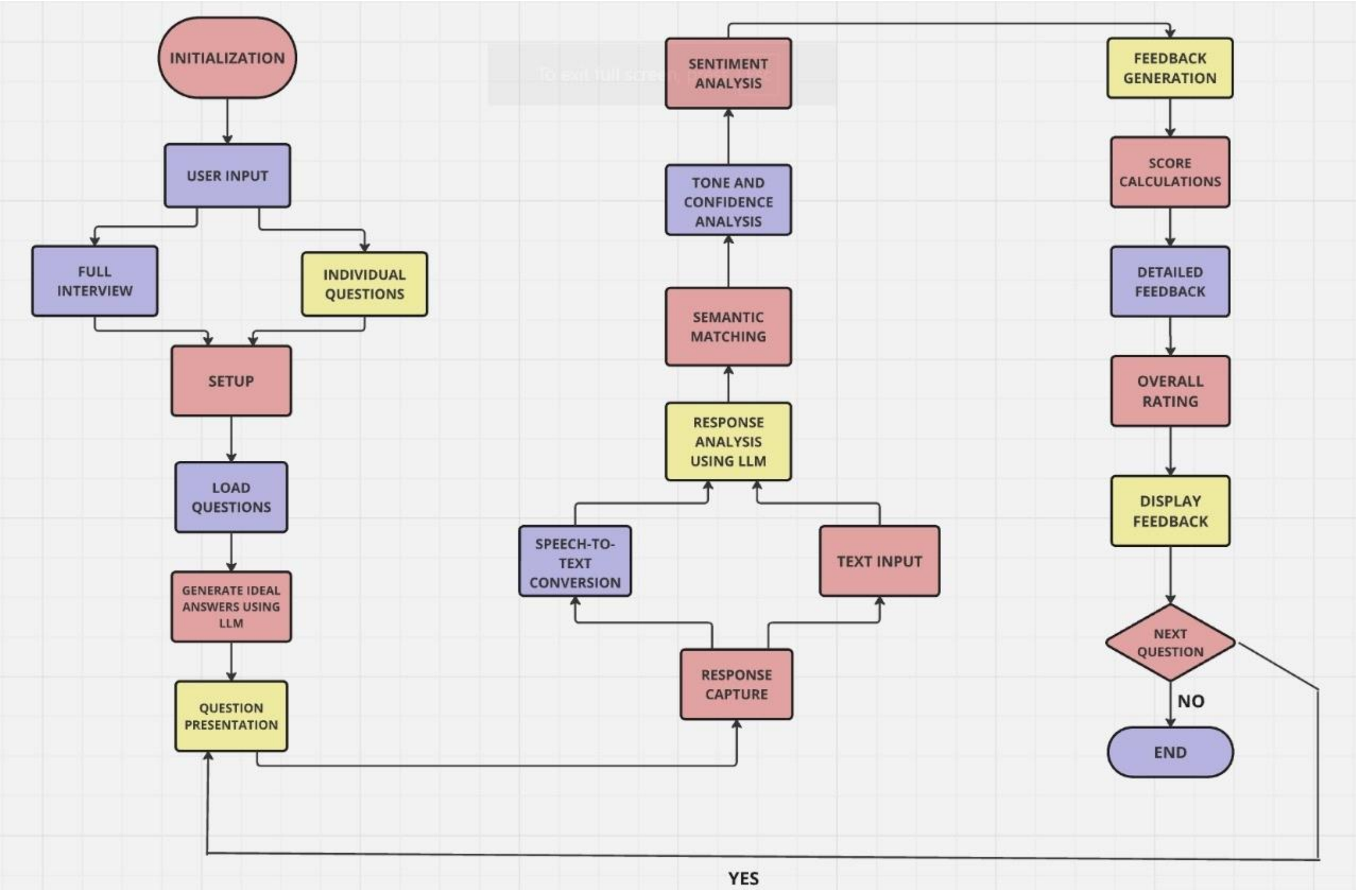

Fig. 1. Block Diagram representing Methodology

maintaining low-latency processing. Continuous updates and learning mechanisms refine the AI-generated feedback, ensuring that the Smart Interview Coach remains adaptive and progressively enhances the candidate's performance over time.

V. RESULTS

The Smart Interview Coach provides a structured evaluation of a candidate's interview performance by analyzing multiple communication attributes. The system assesses pitch, loud- ness, pause duration, filler words, sentiment, and semantic similarity, assigning each attribute a numerical score. These individual scores are then combined using a weighted formula to calculate an overall performance score, ensuring a balanced evaluation that reflects both verbal and linguistic aspects of communication.

The attribute scores obtained from a sample evaluation are presented in Table 1. The scoring system helps assess vocal clarity, fluency, emotional tone, and relevance of responses. Attributes such as loudness and filler word usage contribute significantly to the overall clarity of speech, while sentiment and semantic similarity indicate how well the response aligns with expected answers. Lower scores in certain attributes suggest areas where improvements can be made to enhance engagement and expressiveness.

Attribute	Score
Pitch	5.35
Loudness	9.2
Pause Duration	8.33
Filler Words	9.0
Sentiment	7.67
Semantic	8.1

Similarity	
Overall Score	8.0

TABLE I ATTRIBUTE SCORES

Using a predefined weighting system, the system computes an overall score that provides a comprehensive assessment of the candidate's interview readiness. The structured feedback generated by the system not only helps candidates identify areas for improvement but also offers actionable insights to enhance verbal communication skills.

The Smart Interview Coach provides textual feedback alongside numerical scores, offering personalized suggestions and highlighting areas for improvement as shown in Fig 2. It analyzes speech patterns, sentiment, and fluency to identify strengths and weaknesses. If any attribute is lacking, the sys- tem suggests various steps to improve. This ensures candidates receive both quantitative evaluation and actionable insights for effective interview preparation.

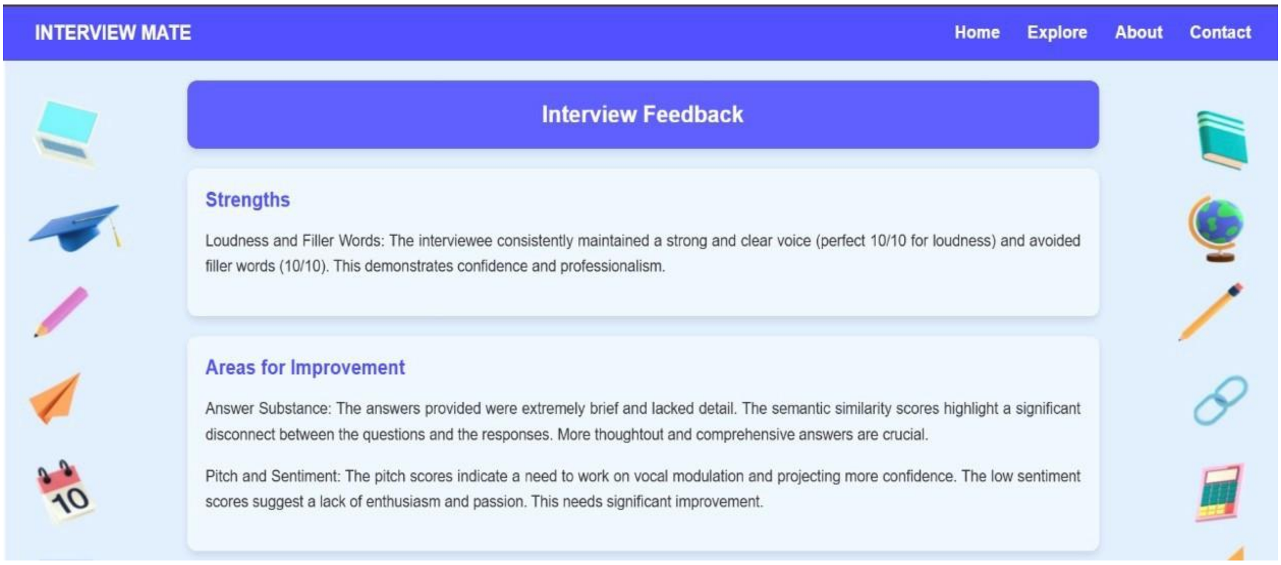

Fig. 2. Feedback Display Page

VI. ANALYSIS OF MODEL PERFORMANCE

Model performance is crucial in tasks like sentiment analy- sis and large language model (LLM) applications, with accu- racy being a key metric. Accuracy indicates how well a model can correctly classify or predict outcomes based on input data. The choice of model depends on factors like task requirements, training data, and whether high precision or domain-specific tuning is needed. While some models are better for general tasks, others may perform better in specialized areas. This analysis underscores the importance of selecting the right model to meet the specific needs of the project.

A. Large Language Model

In comparing LLMs for accuracy, Gemini Flash 1.5 outper- forms other models, achieving the highest accuracy of 96.02% and a sensitivity rate of 94.63% as presented in Table II. This makes it the most reliable model for tasks requiring precise predictions and strong pattern recognition. Its balance of high accuracy and sensitivity gives it an edge in complex applications where both prediction and pattern identification are critical.

While GPT-4 follows closely with an accuracy of 95.00% and a sensitivity of 95.97%, it still lags behind Gemini Flash

1.5 in overall accuracy. Claude Sonnet 3.5, with an accuracy of 94.69%, shows good performance but has a notably lower sensitivity rate of 88.59%, making it less effective in tasks that demand high identification precision. Therefore, Gemini Flash 1.5 remains the most robust model for ensuring both high accuracy and reliable sensitivity.

This makes Gemini Flash 1.5 ideal for tasks requiring real-time accuracy, such as interview analysis. Its balanced sensitivity ensures reliable classifications, reducing errors. Compared to other models, it offers a more consistent and precise performance, making it the preferred choice for AI- driven applications.

Model	Accuracy Rate	Sensitivity Rate
Gemini Flash 1.5	**96.02%**	**94.63%**
GPT-4	95.00%	95.97%
Claude Sonnet 3.5	94.69%	88.59%

TABLE II
MODEL ACCURACY AND SENSITIVITY RATES

B. Sentiment Analysis Model

The RoBERTa model demonstrates strong performance, achieving 93-95% accuracy when allowing for one-star variation. It is particularly advantageous for multilingual sentiment analysis, making it ideal for global or diverse datasets. Its flexibility allows it to handle minor sentiment variations, which is valuable for applications like interview analysis where slight discrepancies in sentiment classification are acceptable.

On the other hand, the CardiffNLP Model offers a lower accuracy of around 67%, making it less suitable for tasks that require precise sentiment detection. Similarly, the Lexa- lytics General Sentiment Model provides 70.5% accuracy, but requires domain-specific tuning to achieve higher accuracy, which may limit its applicability for broader use cases. This is neatly presented in Table III.

Model	Accuracy
RoBERTa Model	**93-95% (one-star variation)**
CardiffNLP Model	67%
Lexalytics General Sentiment Model	up to 80% (domain-specific tuning)

TABLE III
MODEL ACCURACY COMPARISON

VII. APPLICATIONS

The Smart Interview Coach has diverse applications across job preparation, HR, corporate training, and language refinement. For job seekers, it provides real-time feedback on speech clarity, filler words, and emotional tone, helping them refine their communication skills and boost confidence. HR teams can leverage it for objective candidate evaluations, reducing bias and streamlining recruitment. Professionals benefit from soft skills training, improving articulation and persuasion in presentations, client interactions, and negotiations. Universities and training institutes can integrate it into career development programs, offering AI-powered coaching to students and early- career professionals.

Additionally, the system aids non-native speakers in ac- cent neutralization and pronunciation correction, improving fluency and clarity in professional settings. It also enhances mock interview platforms and corporate training by refining customer-facing communication, ensuring employees deliver clear and confident responses. By providing AI-driven, real- time feedback, the Smart Interview Coach revolutionizes in- terview preparation and professional communication, making it more accessible, objective, and effective across multiple domains.

VIII. CONCLUSION AND FUTURE SCOPE

The Smart Interview Coach is an AI-driven platform de- signed to help candidates improve their interview performance through personalized feedback. It analyzes speech clarity, sentiment, and filler word usage, offering actionable insights to enhance communication skills. Unlike traditional interview preparation methods, this system provides objective, data-driven feedback to boost candidates' confidence and readiness. With its potential for expansion in real-time interaction, multi- language support, and corporate training, the Smart Interview Coach aims to revolutionize interview preparation and profes- sional skill development.

The Smart Interview Coach has significant potential for future enhancements. Real-time feedback integration can allow candidates to adjust their responses instantly. Expanding multi-language support will make the system accessible to a global audience, helping non-native speakers improve their communication skills. Advanced emotion detection, including microexpression and vocal stress analysis, can enhance the evaluation process. The system can also be integrated into cor- porate recruitment and educational institutions for large-scale candidate assessments. Additionally, domain-specific feedback tailored for various industries and interactive progress-tracking features can further refine the learning experience, making it a comprehensive tool for career development.

REFERENCES

[1] F. Nagasawa, S. Okada, T. Ishihara, and K. Nitta, ”Adaptive Interview Strategy Based on Interviewees' Speaking Willingness Recognition for Interview Robots,” IEEE Transactions on Affective Computing, vol. 15, no. 3, 2024.

[2] S. Hosseini, J. Quan, X. Deng, T. Nozawa, and Y. Miyake, ”Avatar- Based Feedback in Job Interview Training Impacts Action Identities and Anxiety,” IEEE Transactions on Affective Computing, vol. 15, no. 3, 2024.

[3] N. Dhariwal, S. C. Akunuri, Shivama, and K. S. Banu, ”Audio and Text Sentiment Analysis of Radio Broadcasts,” IEEE Access, vol. 11, 2023.

[4] V. G. R. Vijay, M. C. Manisha, M. D. Mounika, N. A. G. Navyashree, and N. K. Nayana, ”Automation of HR Interview using Deep Learning and NLP,” International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 3, no. 7, 2023.

[5] S. Gupta and Y. Chen, ”Supporting Inclusive Learning Using Chatbots? A Chatbot-Led Interview Study,” Journal of Information Systems Edu- cation, vol. 33, no. 1, pp. 98-108, 2022.

[6] M. Omar, S. Choi, D. Nyang and D. Mohaisen, ”Robust Natural Lan- guage Processing: Recent Advances, Challenges, and Future Directions,” IEEE Access, vol. 10, pp. 86038-86056, 2022.

[7] J. Siswanto, S. Suakanto, M. Andriani, M. Hardiyanti, and T. F. Kusumasari, ”Interview Bot Development with Natural Language Pro- cessing and Machine Learning,” Int. J. Technol., vol. 13, no. 2, pp. 274-285, 2021.

[8] A. A. Shoukat Ali and V. K. Shandilya, ”AI-Natural Language Process- ing (NLP),” Int. J. Res. Appl. Sci. Eng. Technol., vol. 9, no. 8, pp. 135, 2021.

[9] S. Seo, S. Na, and J. Kim, ”HMTL: Heterogeneous Modality Transfer Learning for Audio-Visual Sentiment Analysis,” IEEE Access, vol. 8,pp. 106260-106271, 2020.

[10] P. Kanna, M. G. Sethuraman, P. Shanmugavadivu, and others, ”An Automated System for the Assessment of Interview Performance through Audio & Emotion Cues,” Proc. ICACCS, 2019.

[11] H.-Y. Suen, K.-E. Hung, and C.-L. Lin, ”TensorFlow-Based Automatic Personality Recognition Used in Asynchronous Video Interviews,” IEEE Access, vol. 7, 2019.

[12] D. Kawade and K. Oza, ”Sentiment Analysis: Machine Learning Ap- proach,” Int. J. Eng. Technol., vol. 9, no. 3, pp. 151-156, 2017.

[13] https://www.interviewbit.com/hr-interview-questions/

[14] https://librosa.org/doc/latest/index.html

[15] https://huggingface.co/roberta-base

[16] https://ai.google.dev/api?lang=python