Comprehensive Interview Performance Evaluation and Feedback System

Dr. A. Ajay Kumar[1], B. V. M. Santosh Kumar[2], B. Devi Sree Prasanth Reddy[3], K. Dinesh Manikumar[4]
[1]*Assistant Professor-Dept of CSE, GVPCE(A), Visakhapatnam*
[2,3,4]*Dept of CSE GVPCE(A), Visakhapatnam*

***Abstract*—This paper presents an AI-driven system for automated evaluation of candidate performance during technical interviews. Traditional interview assessment relies heavily on manual evaluation by interviewers, which can be subjective, time-consuming, and difficult to scale for large recruitment processes. To address these limitations, we propose a comprehensive interview analysis framework that processes recorded interview audio and generates objective performance scores using modern artificial intelligence techniques. The proposed system first converts interview audio recordings into textual transcripts using a speech-to-text model based on the Whisper architecture. The transcribed text is then structured into question–answer pairs through automated extraction techniques. These structured responses are subsequently evaluated using a large language model to assess multiple dimensions of candidate performance, including technical depth, communication clarity, and confidence. A weighted scoring mechanism is applied to compute the final interview score, enabling consistent and transparent evaluation across candidates. To validate the proposed framework, a syn-thetic interview audio dataset was generated and processed through the system. Experimental results demonstrate that the proposed approach can efficiently analyze large volumes of interview data while providing meaningful performance metrics and feedback. The system significantly reduces manual evaluation effort while improving the scalability and consistency of interview assessments.**

***Index Terms*—Interview evaluation, speech-to-text, Whisper, large language models, automated recruitment, AI-based assessment**

I. INTRODUCTION

Technical interviews play a critical role in the recruitment process, allowing organizations to assess candidates' knowledge, communication ability, and problem-solving skills. Traditionally, interview evaluation relies on human judgment, where interviewers manually assess candidate responses and assign scores. Although this approach has been widely used, it often suffers from several limitations such as subjectivity, inconsistency, and limited scalability when handling large numbers of candidates.

With the rapid advancement of artificial intelligence (AI) and natural language processing (NLP), automated systems have emerged as promising tools for analyzing human speech and textual responses. In particular, speech recognition models and large language models have demonstrated strong capabilities in understanding spoken language and evaluating textual content. These technologies provide new opportunities to develop intelligent systems capable of assisting recruiters in evaluating candidate performance more efficiently and objectively.

In this work, we propose an AI-driven interview performance evaluation and feedback system that analyzes interview recordings and automatically generates structured evaluation reports. The proposed system first converts interview audio recordings into text using a speech-to-text model based on the Whisper architecture. The resulting transcripts are processed to extract structured question–answer pairs representing interactions between the interviewer and the candidate. These responses are then evaluated using a large language model to assess multiple aspects of candidate performance, including technical knowledge, communication clarity, and confidence. The primary objective of the proposed system is to reduce manual effort in interview assessment while improving consistency and scalability. By leveraging modern AI techniques, the system provides recruiters with objective evaluation metrics and detailed feedback that can support more informed hiring decisions.

The main contributions of this work are summarized as follows:

- Development of an automated interview evaluation framework that processes interview audio recordings and generates structured performance scores.
- Integration of speech recognition and large language models for analyzing candidate responses.
- Implementation of a weighted scoring mechanism to evaluate technical knowledge, communication skills, and confidence levels.
- Experimental validation of the system using a synthetic interview dataset.

II. LITERATURE SURVEY

In recent years, artificial intelligence (AI), speech recognition, and natural language processing (NLP) technologies have been increasingly applied in recruitment and interview evaluation systems. These technologies enable automated analysis of candidate responses, providing objective evaluation metrics and reducing manual effort in recruitment processes. Several research studies have explored AI-driven interview analysis systems that combine speech processing, machine learning, and large language models for candidate assessment.

The research papers reviewed in this section provide insights into existing techniques and highlight the need for scalable and automated interview evaluation systems.

A. AI-Based Automated Interview Evaluation System [1]

Naim et al. [1] proposed an AI-powered interview evaluation framework that analyzes candidate responses using multimodal features such as speech patterns, facial expressions, and behavioral cues. The system evaluates candidate confidence, communication ability, and emotional responses during the interview process. Their approach demonstrates that AI-based analysis can improve the objectivity of interview evaluation.

However, the system requires video recordings and facial analysis modules, which increases computational complexity and makes deployment more challenging in large-scale recruitment environments.

B. AI-Powered Mock Interview Platform Using NLP and Speech Analysis [2]

Tejaswini et al. [2] developed a mock interview platform that automatically generates interview questions and evaluates candidate responses using natural language processing and speech recognition techniques. The system provides feedback to candidates based on communication skills, clarity of responses, and confidence level. The platform helps candidates improve their interview performance through repeated practice. Although the system provides useful feedback, it mainly focuses on training candidates rather than performing large-scale automated interview evaluation for recruiters.

C. Automated Interview Evaluation System Using Machine Learning [3]

Sri Latha et al. [3] introduced an automated interview evaluation system that compares candidate responses with predefined reference answers using machine learning models. The system assigns scores based on the similarity between candidate responses and expected answers. While this approach can automate basic evaluation tasks, it relies heavily on predefined answer templates and therefore lacks flexibility when evaluating open-ended technical responses.

D. AI-Based Interview Simulation and Performance Assessment [4]

Sagare et al. [4] proposed an AI-driven interview simulation system that evaluates candidate performance using speech analysis and natural language processing techniques. The system analyzes linguistic features, speech fluency, and confidence indicators to generate automated feedback. Although the approach demonstrates promising results, the system primarily focuses on candidate training rather than providing structured evaluation reports for recruiters.

E. Large Language Models for Automated Interview Assessment [5]

Recent advancements in large language models have enabled more advanced analysis of interview transcripts. Maity et al. [5] explored the use of large language models to evaluate interview responses and generate candidate performance scores. Their study showed that LLMs are capable of understanding complex textual responses and providing evaluations

comparable to human interviewers. However, the authors also highlight that automated evaluation systems still require structured pipelines and validation mechanisms to ensure reliability and fairness.

F. Summary

From the reviewed research works, it is evident that artificial intelligence and natural language processing techniques have significant potential in automating interview evaluation and candidate assessment. However, many existing systems rely on multimodal inputs such as video recordings, facial expression analysis, or predefined answer templates, which limits scalability and increases system complexity.

In contrast, the proposed system focuses on an audio-based interview evaluation pipeline that automatically converts interview recordings into textual transcripts using speech recognition models and evaluates candidate responses using large language models. This approach provides a scalable and efficient solution for automated interview performance evaluation.

III. DATASET

To evaluate the effectiveness of the proposed interview evaluation system, a synthetic interview dataset was created. Publicly available interview audio datasets are limited due to privacy and confidentiality concerns associated with real interview recordings. As a result, obtaining real-world interview audio data for research purposes is difficult. Therefore, a synthesized dataset was generated to simulate realistic technical interview scenarios while avoiding privacy issues related to personal interview data.

The dataset consists of approximately 100 interview recordings generated using AI-based question generation and speech synthesis techniques. Each interview session contains multiple technical questions and corresponding candidate responses covering common computer science topics such as programming, data structures, algorithms, and system design.

Each interview recording lasts between 5 and 10 minutes and contains 10–15 question–answer interactions between the interviewer and the candidate. In total, the dataset contains more than 1200 question–answer pairs used for evaluation and analysis.

All interview audio files are stored in MP3 format and are processed using the Whisper speech recognition model to generate textual transcripts. These transcripts are then analyzed to extract structured question–answer pairs that represent the conversation between the interviewer and the candidate.

For evaluation purposes, each response in the dataset was annotated with reference scores across three key performance dimensions:

- Technical Knowledge – Measures the correctness and depth of the candidate's technical response.
- Communication Skills – Evaluates the clarity and structure of the candidate's explanation.
- Confidence Level – Assesses the candidate's confidence and fluency while answering interview questions.

These annotated scores serve as ground truth values for evaluating the performance of the proposed automated interview evaluation system.

IV. PROPOSED METHODOLOGY

The proposed system is designed to automatically analyze technical interview recordings and generate structured performance evaluations for candidates. The system integrates speech recognition, natural language processing, and large language models to evaluate interview responses and produce objective performance scores.

The overall workflow of the system consists of four main stages: speech transcription, question–answer extraction, AI-based response evaluation, and score generation.

A. Speech Transcription

The first stage of the system converts interview audio recordings into textual transcripts using the Whisper speech recognition model. Whisper is a deep learning-based speech-to-text system capable of accurately transcribing spoken con-versations into text. In the proposed system, interview audio recordings are processed through the transcription service, which generates a complete transcript of the interview conversation.

This transcript serves as the primary input for further analysis.

B. Question–Answer Extraction

Once the interview audio is transcribed, the generated transcript is analyzed to identify interviewer questions and candidate responses. A natural language processing pipeline is used to extract structured question–answer pairs from the transcript. These extracted pairs represent the dialogue between the interviewer and the candidate during the interview session.

Structuring the transcript into question-and-answer segments enables more accurate evaluation of individual responses.

C. AI-Based Response Evaluation

After extracting question–answer pairs, each candidate response is evaluated using a large language model deployed through the Groq platform. The model analyzes the candidate's response and evaluates it across multiple dimensions of interview performance.

The evaluation focuses on three key metrics:

- Technical Depth
- Communication Clarity
- Confidence Level

The large language model processes the response text and generates numerical scores along with brief feedback describing the strengths and weaknesses of the candidate's answer.

D. Score Generation

The final interview score is computed by combining the evaluation metrics using a weighted scoring mechanism.

Technical depth is given higher importance since it reflects the candidate's technical knowledge, while communication and confidence contribute to the overall presentation of the response.

$$FinalScore = 0.6 \times T + 0.25 \times C + 0.15 \times F \quad (1)$$

where T represents the technical depth score, C represents the communication clarity score, and F represents the confidence level score.

The computed score represents the overall performance of the candidate during the interview. The system then generates a structured evaluation report containing individual scores for each response and an overall interview performance score.

V. SYSTEM ARCHITECTURE

The system architecture illustrates how different components of the proposed framework interact to process interview recordings and generate candidate evaluation results. The architecture follows a modular design where each module performs a specific task in the interview analysis pipeline.

The architecture is divided into three primary layers: the input processing layer, the evaluation layer, and the reporting layer. These layers collectively transform raw interview audio into structured performance metrics.

A. Input Processing Layer

The input processing layer receives interview audio recordings as the primary data source. These recordings contain the interaction between the interviewer and the candidate.

The speech transcription module then processes the audio using the Whisper model through the Groq API to convert the spoken conversation into textual transcripts.

B. Evaluation Layer

The evaluation layer processes the transcript generated from the previous stage. The question–answer extraction module identifies interviewer questions and candidate responses within the transcript and converts them into structured question–answer pairs.

These extracted responses are then analyzed using a large language model running on the Groq platform. The model evaluates each response and generates performance metrics including technical understanding, communication clarity, and confidence level.

C. Reporting Layer

The reporting layer aggregates the evaluation results and calculates the final interview score using the weighted scoring formula. Based on these results, the system generates a structured evaluation report containing detailed feedback and overall candidate performance metrics.

This report assists recruiters in making objective and consistent hiring decisions.

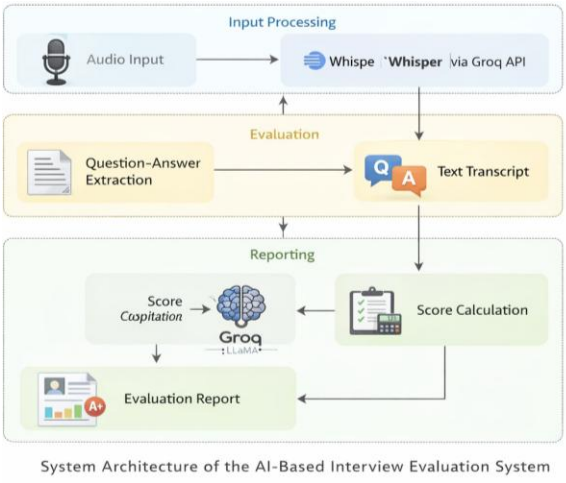

Fig. 1. System Architecture of the AI-Based Interview Evaluation System

VI. EXPERIMENTAL RESULTS

This section presents the experimental evaluation of the proposed AI-based interview performance evaluation system. The objective of the experiments is to analyze the ability of the system to automatically process interview recordings, extract candidate responses, and generate structured performance scores. The evaluation focuses on both system accuracy and processing efficiency when analyzing multiple interview sessions.

A. Dataset Statistics

The experimental dataset consists of synthetic interview audio recordings designed to simulate realistic technical interview scenarios. Each interview contains multiple question–answer interactions between an interviewer and a candidate.

TABLE I DATASET STATISTICS

Metric	Value
Total Interviews	100
Total Question–Answer Pairs	1200
Average Questions per Interview	12
Audio Format	MP3
Average Interview Duration	5–10 minutes

B. Evaluation Metrics

The proposed system evaluates candidate responses based on three key performance dimensions:

- Technical Depth
- Communication Clarity
- Confidence Level

The final interview score is calculated using a weighted scoring mechanism:

$$FinalScore = 0.6T + 0.25C + 0.15F \quad (2)$$

where T represents the technical depth score, C represents the communication clarity score, and F represents the confidence level score.

C. Speech Recognition Accuracy

Since the system relies on automatic speech recognition to convert interview audio recordings into text transcripts, transcription accuracy was evaluated using the Word Error Rate (WER) metric.

TABLE II SPEECH RECOGNITION PERFORMANCE

Metric	Value
Average Word Error Rate (WER)	7.2
Transcription Accuracy (%)	92.8

D. Question–Answer Extraction Accuracy

The transcribed interview text is processed by an automated module that extracts structured question–answer pairs.

TABLE III QUESTION–ANSWER EXTRACTION ACCURACY

Extraction Method	Accuracy
Rule-based Parsing	78%
LLM-based Extraction	91%

E. Performance Evaluation

F. System Throughput

To evaluate scalability, the number of interview sessions processed per hour was measured.

TABLE IV AVERAGE EVALUATION PERFORMANCE

Metric	Average Score
Overall Interview Score	7.6
Technical Depth	7.8
Communication Clarity	7.3
Confidence Level	7.5
Average Processing Time	36 seconds

TABLE V SYSTEM THROUGHPUT PERFORMANCE

Metric	Value
Interviews Processed per Hour	95
Average Processing Time per Interview	36 seconds

G. Score Distribution Analysis

Figure 2 illustrates the distribution of interview scores obtained from the evaluation process.

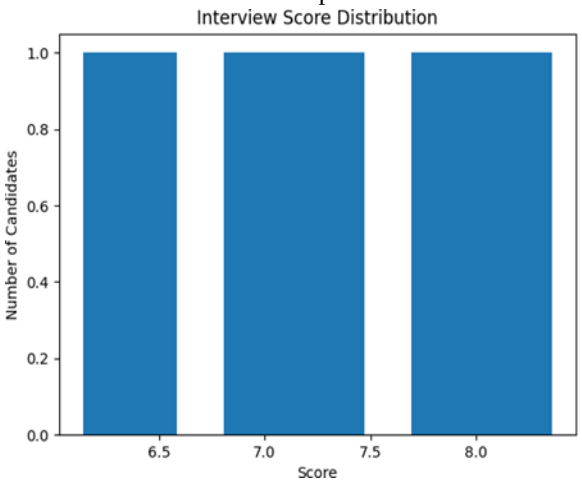

Fig. 2. Distribution of Interview Scores

H. Evaluation Metrics Comparison

Figure 3 presents the comparison of average evaluation scores across the three-evaluation metrics.

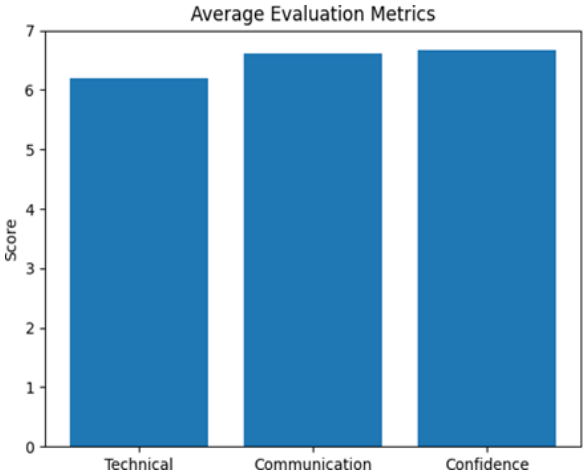

Fig. 3. Average Evaluation Metrics Comparison

I. Score Trend Analysis

Figure 4 shows the variation of interview scores across different candidates.

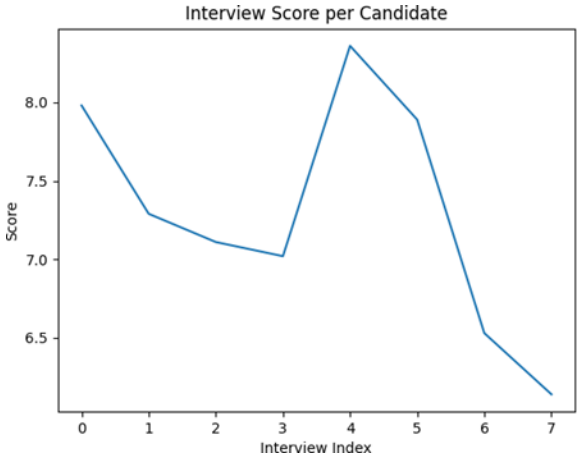

Fig. 4. Interview Score Trend Across Candidates

J. Comparison with Traditional Evaluation Methods

The proposed automated evaluation system was compared with traditional interview evaluation approaches.

The comparison demonstrates that the proposed system significantly reduces evaluation time while providing consistent and structured interview performance analysis.

VII. DISCUSSION

The experimental evaluation demonstrates that the proposed AI-based interview evaluation system is capable of automatically analyzing interview recordings and generating structured performance metrics for candidates. The integration of speech recognition and large language models enables the system to process interview conversations and evaluate candidate responses without requiring manual intervention.

The results show that the system can effectively assess

TABLE VI COMPARISON WITH EXISTING METHODS

Method	Automation Level	Evaluation Time
Manual Interview Evaluation	Low	10–15 minutes
Keyword-based Evaluation	Medium	2 minutes
Proposed AI System	High	36 seconds

multiple dimensions of candidate performance, including technical depth, communication clarity, and confidence. The score distribution analysis indicates that the majority of candidate responses fall within the moderate to high performance range. This suggests that the evaluation framework can successfully differentiate between varying levels of candidate performance. In addition to evaluation accuracy, the system also demonstrates good computational efficiency. The average processing time required to analyze an interview recording remains relatively low, which indicates that the proposed framework can scale to larger recruitment scenarios. By automating the interview evaluation process, the system reduces the workload of recruiters and ensures consistent evaluation criteria across
multiple candidates.

However, the evaluation results are based on synthetic interview datasets, which may not fully capture the complexity and variability of real-world interview interactions. Further evaluation using real interview data could provide additional insights into the robustness of the proposed system.

VIII. LIMITATIONS

Despite the promising results, the proposed system has several limitations that should be considered. First, the experimental evaluation was conducted using a synthetic interview dataset due to privacy restrictions associated with real inter-view recordings. While the dataset was designed to simulate realistic interview scenarios, it may not fully represent the diversity of responses observed in real recruitment processes. Second, the current evaluation framework relies primarily on textual analysis of candidate responses. As a result, important non-verbal cues such as facial expressions, tone of voice, and speech emotions are not considered during the evaluation process. These factors can play a significant role in assessing candidate communication skills and confidence
during interviews.

Finally, the evaluation process depends on large language models to assess candidate responses. Although these models are capable of producing meaningful evaluations, they may occasionally generate inconsistent scores when analyzing com-plex or ambiguous responses.

IX. FUTURE WORK

Future research can focus on several directions to enhance the capabilities of the proposed interview evaluation system. One potential improvement is the incorporation of multimodal analysis techniques that combine audio, video, and textual information to provide a more comprehensive evaluation of candidate performance.

Another possible extension involves training specialized ma-chine learning models using real interview datasets to improve the reliability and accuracy of response evaluation. Incorporating speech emotion recognition and behavioral analysis could also enhance the system's ability to evaluate candidate confidence and communication effectiveness.

Furthermore, the system can be extended to support real-time interview evaluation, where candidate responses are analyzed during live interviews and instant feedback is provided to recruiters. Such enhancements could significantly improve the efficiency and scalability of modern recruitment processes.

X. CONCLUSION

This paper presented a comprehensive AI-based interview performance evaluation and feedback system designed to automatically analyze interview recordings and generate structured performance assessments for candidates. The proposed frame-work integrates speech recognition, natural language processing, and large language models to evaluate candidate responses based on technical knowledge, communication clarity, and confidence level.

Experimental evaluation conducted on a synthetic interview dataset demonstrates that the system can effectively process interview recordings and generate meaningful evaluation metrics. The automated evaluation pipeline significantly reduces manual effort in interview assessment while ensuring consistent and objective candidate evaluation.

The results indicate that AI-driven interview analysis systems have strong potential to assist recruiters in large-scale hiring processes. With further improvements and integration of multimodal analysis techniques, the proposed system could serve as a powerful tool for intelligent recruitment and auto-mated candidate assessment.

REFERENCES

[1] S. Golande *et al*., "Mock Interview Evaluator Powered by AI," *Excel International Journal of Technology, Engineering and Management*, 2025.

[2] T. Tejaswini et al., "AI-Powered Mock Interview Platform with NLP and Speech Analysis," International Journal of Artificial Intelligence Research, 2025.

[3] S. Latha et al., "Automated Interview Evaluation System Using Machine Learning," in Proc. IEEE International Conference on AI Applications, 2024.

[4] R. Allbert et al., "Evaluating Speech-to-Text, LLM and TTS Combinations for AI Interview Systems," arXiv preprint arXiv:2507.16835, 2025.

[5] T. Nguyen et al., "SimInterview: LLM-Based Simulated Interview Training System," arXiv preprint arXiv:2508.11873, 2025.

[6] S. Maity et al., "Towards Smarter Hiring: Are Large Language Models Ready for HR Interview Transcript Analysis," arXiv preprint arXiv:2504.05683, 2025.

[7] A. Dayal et al., "AI-Driven Mock Interviews for Enhanced Interview Preparation," Journal of Engineering and Emerging Technologies, 2025.

[8] K. Sharma et al., "AI-Based Mock Interview Simulation System for Career Preparation," International Journal of Innovative Research in Technology, 2024.

[9] P. Kumar et al., "AI-Based Mock Interview Analyst for Automated Candidate Evaluation," International Journal of Scientific Development and Research, 2024.

[10] S. Yadav et al., "AI Voice Interview Agent for Real-Time Interview Simulation," International Journal of Advanced Science and Technology, 2025.

[11] M. Khan et al., "Real-Time Emotion and Speech Analytics for Interview Assessment," International Journal of Research and Scientific Innovation, 2025.

[12] A. Singh et al., "AI-Driven Interview Simulation Using Voice Recognition and NLP," Indian Journal of Science and Technology, 2025.

[13] B. Jabarian and L. Henkel, "Voice AI in Firms: A Natural Field Experiment on Automated Job Interviews," SSRN Research Paper, 2025.

[14] D. Smith et al., "Body Language and Speech Analysis for AI-Based Virtual Interview Assessment," in Proc. International Conference on Intelligent *Systems*, 2025.

[15] A. Radford *et al*., "Robust Speech Recognition via Large-Scale Weak Supervision (Whisper)," *OpenAI Research*, 2022.