Web Based Selector – Applicant Simulation Software

Mr. Rupesh Dhadwe[1], Mr. Abhishek More[2], Mr. Nilesh Darade[3], Mr. Abhishek Shingokar[4], Ms. Kanchan Pekhale[5]

Met's Institute of Engineering Savitribai Phule Pune University

Abstract

This project develops a "Web Based Selector – Applicant Simulation Software" that helps users prepare for real-world job interviews through domain-specific mock sessions. Users can register, log in, and choose from technical domains like Machine Learning, Java, and Android. The system presents 10 domain-specific questions, and users respond via audio while machine learning algorithms analyze their facial expressions and voice tone. Facial analysis detects emotions such as confidence and nervousness by examining features like smiles and eye contact, while voice analysis evaluates speech rate, pitch, and tone variations. After the interview, the system generates a detailed performance report with feedback on verbal responses, non-verbal cues, and confidence levels, along with suggestions for improvement to enhance interview skills.

Keywords: Smart Interview System, Mock Interviews, Domain-specific Questions, Facial Expression Analysis, Voice Tone Analysis, Machine Learning

INTRODUCTION

In today's highly competitive job market, succeeding in interviews is a critical step toward securing employment. An interview is not only a test of technical knowledge but also a reflection of a candidate's communication skills, confidence, and emotional intelligence. However, traditional approaches to interview preparation often emphasize

technical proficiency while neglecting the critical non-verbal aspects of communication. These overlooked areas—such as body language, facial expressions, and voice modulation—can significantly influence an interviewer's perception and ultimately determine a candidate's success.

Recognizing these challenges, we propose a Web-Based Selector–Applicant Simulation Software, designed to bridge the gap in interview preparation. The system integrates advanced machine learning techniques to create a platform where candidates can practice domain-specific mock interviews while receiving detailed, real-time feedback on their performance. This innovative solution caters to both technical and behavioral aspects of interview readiness, ensuring holistic preparation for real-world scenarios.

The proposed system provides a user-friendly interface for candidates to register, log in, and participate in mock interviews tailored to specific domains such as Machine Learning, Java, and Android. Users answer a set of domain-specific questions via audio responses. During these sessions, the system captures and analyzes the candidates' verbal and non-verbal cues. Advanced machine learning algorithms assess speech content for accuracy and coherence while evaluating facial expressions and vocal tone for indicators of confidence, attentiveness, and emotional state.

The facial expression analysis module detects subtle emotional cues such as smiles, frowns, and eye movements, which are critical indicators of confidence and engagement. Similarly, the voice tone analysis component evaluates speech rate, pitch, and tonal variation to assess stress levels, confidence, and overall emotional stability. By leveraging these insights, the system provides candidates with actionable feedback and personalized suggestions for improvement, thereby addressing the nuanced aspects of interview preparation.

This comprehensive approach to mock interviews not only enhances technical proficiency but also builds the soft skills that are indispensable in professional settings. The system aims to empower candidates by fostering self-awareness and equipping them with the tools to excel in high-pressure interview environments.

PROBLEM STATEMENT

The primary problem faced by candidates during interview preparation is the lack of a comprehensive system that evaluates both technical knowledge and non-verbal communication skills. While most candidates focus heavily on answering technical questions, they often neglect important soft skills such as facial expressions, voice tone, and confidence. This results in poor performance during real interviews, even when they have the right answers. Current interview preparation methods and platforms mainly assess technical abilities, without providing any feedback on the non-verbal cues that influence interview outcomes. Consequently, candidates are not fully prepared for real interviews, where both technical expertise and communication skills are crucial for success. There is a clear need for an integrated system that can test and provide feedback on both technical and behavioral aspects of an interview, helping candidates improve holistically.

OBJECTIVE

1. Provide Domain-Specific Mock Interviews Allow users to practice mock interviews in various domains such as Machine Learning, Java, Android, etc.

2. Analyze Verbal Responses Evaluate users answers to interview questions based on their content, accuracy, and communication skills.

3. Assess Non-Verbal Cues Use facial expression analysis and voice tone assessment to measure users' confidence and emotional state during the interview.

4. Generate Detailed Performance Reports provide a comprehensive report that includes feedback on both verbal and non-verbal communication, confidence levels, and emotional cues.

5. Offer Suggestions for Improvement Provide actionable insights and suggestions to help users improve their interview performance, focusing on areas of weakness.

LITERATURE SURVEY

It is a system for public speaking practice that analyzes speech, facial expressions, and gestures. The system provides valuable feedback to improve confidence and overall performance during public speaking sessions.

It helps individuals refine communication skills by evaluating their presentation in real-time using multimodal inputs.[1]

The System that assesses interview performance through machine learning techniques. It evaluates emotions, engagement, and confidence by analyzing audio and visual data during an interview. By offering detailed feedback, it helps candidates develop deeper self-awareness and identify areas for improvement.[2]

An emotion detection system designed to assist candidates in preparing for virtual interviews. By analyzing facial expressions and vocal cues, the system detects the emotional state of the candidate, offering real-time feedback to reduce interview anxiety and improve response quality, thus enhancing interview readiness. [3]

The combination of voice tone and facial expression analysis for candidate evaluation in virtual interviews. The study highlights how both vocal and facial cues assess a candidate's confidence and communication skills. The feedback provided guides individuals toward better interview performance by pinpointing strengths and areas needing attention. [4]

A Virtual Interview Assistant that uses facial and vocal emotion recognition to assess candidate emotions during interviews. The system provides valuable insights into emotional state, confidence, and interview readiness, thus offering a personalized approach to improving interview performance.[5]

SYSTEM ARCHITECTURE

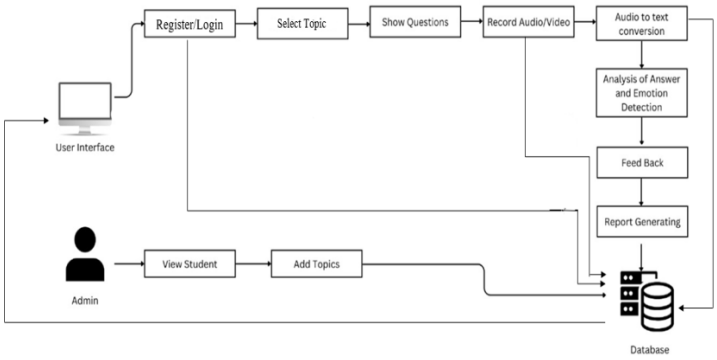

Fig 01: Web Based Selector – Applicant Simulation Software

The Web Based Selector – Applicant Simulation Software uses artificial intelligence (AI) and machine learning to analyze how users respond during a practice interview and provide useful feedback. For facial expression analysis, the system records video using a webcam and uses deep learning models to recognize emotions like confidence, nervousness, and engagement by analyzing facial features such as eye contact, smiles, and eyebrow movements. In voice analysis, the system listens to the user's speech, studying tone, pitch, and speaking speed to understand confidence levels.

For answer evaluation, the system converts spoken answers into text using speech recognition software like Google Speech-to-Text. It then analyzes the response using language processing models to check if the answer is relevant, well-structured, and technically correct. Finally, the system creates a detailed performance report, rating the user's verbal responses, facial expressions, and confidence levels. It also provides personalized feedback and suggestions for improvement. By combining AI-powered speech, facial

analysis, and smart feedback, the system helps users get better at interviews by improving both their technical answers and body language.

The system architecture consists of several interconnected components:

User Interface (UI): The user selects a topic, gives permission, and records audio/video responses. The UI is connected to the database for storing user data and session details.

Recording and Analysis: Audio and video responses are analyzed for emotional tone, content, and facial expressions using AI models.

Feedback and Report Generation: Based on the analysis, personalized feedback is generated and a report is created, which is accessible to the user.

Database (DB): Stores user profiles, media files, analysis results, and feedback.

Admin View: Administrators can add topics, manage users, and monitor progress through a separate interface connected to the database.

METHODOLOGY

Convolutional Neural Networks (CNNs) are employed to process video frames captured during mock interviews. The system detects emotions such as confidence and nervousness by analysing visual features like eye movements, smiles, and facial expressions. The CNN model extracts and classifies these features in real-time, enabling the system to evaluate the user's non-verbal communication and engagement level.

Recurrent Neural Networks (RNNs) are utilized to analyze sequential audio data, assessing voice tone, pitch, and speech patterns. The system leverages RNNs, often combined with Long Short-Term Memory (LSTM) units, to capture temporal dependencies in the user's speech. This analysis helps identify stress, confidence, and emotional consistency throughout the interview, providing valuable insights into vocal delivery.

Gemini API models are integrated for generating domain-specific interview questions dynamically, ensuring relevance and variety. The system uses Cosine similarity algorithm to analyze user responses for technical accuracy, coherence, and content quality. Furthermore, it generates detailed, personalized feedback and actionable suggestions, helping users improve their technical knowledge and communication skills.

ALGORITHM

1. Speech Recognition:

In a web-based applicant simulation software, speech recognition serves as a core technology to enable seamless interaction and real-time response evaluation during mock interviews. SpeechRecognition librarysupports several speech-to-text engines, including Google's Web Speech API, the software transcribes the applicant's spoken responses into text, allowing the system to process and analyze them. This process is facilitated by machine learning models that ensure high accuracy in transcription, even in the presence of various accents, pronunciations, or background noise.

2. Natural Language Processing:

The web-based applicant simulation software leverages advanced Natural Language Processing (NLP) techniques to enhance the interview preparation experience. NLP plays a critical role in generating role-specific interview questions tailored to the user's profile, based on the job description and required skills. During the mock interview, the software uses Speech Recognition, a subfield of NLP, to convert the user's spoken responses into text. Once transcribed, NLP algorithms analyze the content of the answers, checking for relevance, structure, and key points. The Gemini API generates domain-specific questions, while the cosine similarity algorithm evaluates user responses against model answers for relevance and accuracy, ensuring precise feedback.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Cos(θ): Similarity score between the candidate's input and the ideal response.

A·B: Alignment between the candidate's performance and the expected behaviour.

‖A‖ and ‖B‖: Strengths of the candidate's input and the ideal response.

∑ AiBi: Agreement between the candidate's input and the expected response.

√∑ Ai² and √∑ Bi²: Normalized magnitudes of input and expected response for comparison.

3. Facial Expression Recognition:

1. **Input Image Frame Extraction:** Video feed from the candidate's camera is divided into individual image frames.
2. **Preprocessing:** Frames are resized and normalized to ensure uniformity in the input dimensions.
3. **Feature Extraction (Convolutional Layers):** The CNN applies convolution operations to extract features like eye movements, lip curvature, and facial expressions.
4. **Pooling (Down sampling):** Reduces the size of feature maps while retaining important features to make processing faster and more efficient.
5. **Flattening:** Converts the pooled feature maps into a 1D vector to prepare for classification.
6. **Classification (Fully Connected Layers):** The extracted features are analysed to classify facial expressions (e.g., neutral, nervous, confident).Outputis a probability distributionover predefined emotion categories.
7. **Integration with Feedback:** The classified expressions are used to provide feedback on non-verbal cues, such as confidence and emotional presence.

4. Emotion Recognition from Voice Tone:

1. Input Audio Sequence Extraction: The candidate's speech is converted into audio frames and broken into sequences of features (e.g., pitch and tone)
2. Preprocessing: The audio signals are normalized to remove noise and ensure clarity.Text responses (if any) are tokenized into word sequences for analysis.
3. Feature Representation: Audio features (e.g., pitch and tone) or word embeddings (e.g., using Word2Vec or GloVe) are used as input vectors to the RNN.

4. Sequence Learning (RNN Layers): The RNN processes input sequences step by step, capturing patterns in tone, sentiment, or relevance over time.
5. Memory Management (LSTM/GRU): Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) layers are used to handle long-term dependencies, ensuring earlier parts of the answer are considered when evaluating the overall response.
6. Output Analysis: The RNN outputs scores for sentiment (e.g., confident, hesitant).
7. Integration with Feedback: Combines tone and semantic analysis to provide detailed feedback on verbal communication, including confidence level and content quality.

TECHNICAL EXPLANATION

The Mock Interview System is an advanced platform that revolutionizes interview preparation by integrating Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) to evaluate both technical knowledge and non-verbal communication. Users can select domains such as Machine Learning, Deep Learning, and Data Mining, etc. after which the system displays domain-specific questions for the candidate to answer. During the response, the platform analyzes the user's voice tone and facial expressions, providing insights into emotional presence and non-verbal cues. This comprehensive approach helps users understand and improve their confidence, communication skills, and emotional stability. By offering detailed feedback on strengths and areas for improvement, the system enables candidates to enhance their technical expertise and overall performance, preparing them for success in competitive job markets and real-world interviews.

FUCTIONAL REQUIREMENTS

1. The system should allow users to register using their email, social media, or custom credentials.

2. Users should be able to securely log in to access the mock interview platform.

3. Users must be able to choose from multiple domains such as Machine Learning, Java, Android, etc.

4. The system should provide a list of relevant technical questions for each selected domain.

5. The system should present a set of 10 questions.

6. Users should be able to answer these questions using audio responses.

7. The system should use facial recognition to analyze the user’s facial expressions.

8. It should also use voice tone analysis to assess confidence, stress levels, and emotional state.

NON FUCTIONAL REQUIREMENTS

1. Usability: The system should have an intuitive user interface that is easy to navigate for both technical and non-technical users.

2. Performance: The system should process and analyze facial expressions and voice tone in real time with minimal delay. The mock interview should run smoothly, even for large numbers of users.

3. Scalability: The system should be able to scale to accommodate a large number of users without compromising performance.

CONCLUSION

In conclusion, the AI-driven interview simulation system offers a comprehensive solution for candidates looking to enhance their interview skills. By integrating real-time analysis of speech, facial expressions, and answer content, it provides a realistic and holistic evaluation of a user's performance. The post-interview in-depth analysis and personalized feedback ensure that users receive actionable insights, helping them identify areas for improvement and boosting their confidence. This innovative approach prepares candidates not only for technical questions but also for handling non-verbal communication effectively, making them better equipped for real-world interviews.

REFERENCES

1.Batrinca et al., "Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training," ACM Transactions on Interactive Intelligent Systems, 2013.

2. Kapoor and Narayanan, "Automatic Assessment of Interview Performance Using Machine Learning," IEEE Transactions on Affective Computing, 2020.

3. Datta et al., "AI-based Emotion Detection System for Virtual Interview Preparation," International Journal of Computer Applications, 2021.

4. Vishwakarma et al., "Voice Tone and Facial Expression Analysis for Candidate Evaluation in Virtual Interviews," Procedia Computer Science (Elsevier), 2019.

5. Kumar et al., "Virtual Interview Assistant: A Facial and Vocal Emotion Recognition System for Interview Preparation," IEEE Access, 2022s.