e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

AI MOCK INTERVIEW

Faraz Pathan[*1], Swati Ghawate[*2], Huzaif Shaikh[*3], Idris Shaikh[*4], Abdullaha Shaikh[*5]

[*1,2,3,4,5]Computer Department Savitribai Phule Pune University Pune, India.

ABSTRACT

The primary goal of this research is to create an AI-based video interview agent capable of predicting communication skills and personality traits by leveraging a multimodal analysis framework. Interviewing is essential. We address the need for interactive platforms that combine AI-driven technologies with personalized feedback to evaluate communication skills and personality traits. By Investigating the application of machine learning (ML) models, including OpenCV, Wave2Vec, and BERT, we can determine their use in assessing communication skills and personality traits during job interviews. We explore how these ML models can objectively evaluate key interview metrics like candidate confidence, speech accuracy, and non-verbal cues, traditionally judged subjectively. Using a dataset of 500 recorded interviews, various techniques were applied to extract and analyze facial expressions, body language, and speech patterns. Results show that ML-based assessments correlate strongly with human evaluations, with an 85% agreement on confidence measures and 78% on answer correctness. These findings suggest that ML models, such as BERT for natural language processing (NLP) and Wave2Vec for speech recognition, can provide objective insights to complement human judgment in the hiring process, reducing bias and improving hiring outcomes.

I. INTRODUCTION

In today's competitive job market, excelling in interviews is vital for job seekers. However, traditional interview evaluations often rely on subjective human judgments, which can introduce bias and inconsistency. Recent advancements in machine learning (ML), particularly in natural language processing (NLP) and speech recognition, offer promising avenues for improving interview preparation and assessment. The "AI-Based Mock Interview Platform" leverages generative AI, including Wave2Vec for accurate speech recognition and BERT for analyzing the meaning of responses, to simulate mock interviews. These tools allow users to record their responses to AI-generated questions and receive detailed feedback on their communication skills, such as clarity, correctness, and confidence. Additionally, OpenCV is used to analyze non-verbal cues like facial expressions and body language.

This survey addresses the need for interactive platforms that combine AI-driven technologies with personalized feedback to evaluate communication skills and personality traits. The research focuses on developing a system capable of generating relevant interview questions, analyzing speech patterns and semantic correctness using Wave2Vec and BERT, and providing detailed insights based on non-verbal cues through OpenCV. By integrating these ML models, the platform offers a comprehensive mock interview experience that not only simulates real-world conditions but also helps users improve their performance and confidence. The ultimate goal is to reduce bias and subjectivity in interview evaluations, thereby democratizing access to high-quality interview preparation tools.

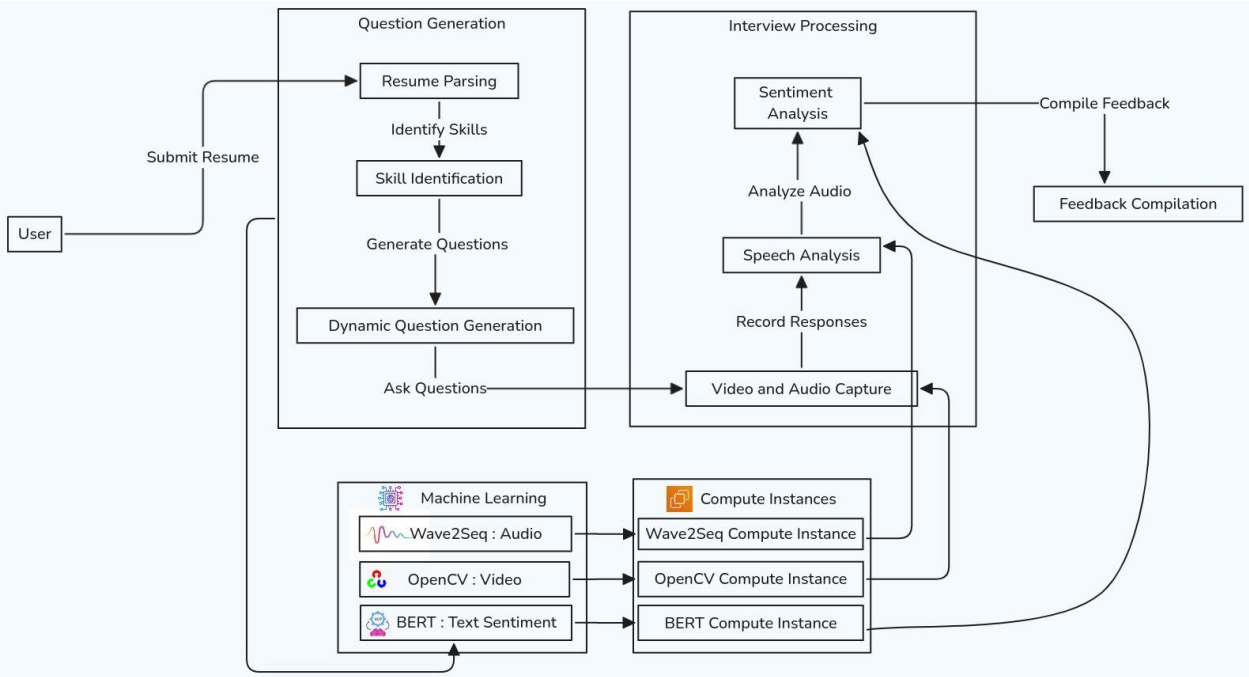
II. METHODOLOGY

The primary goal of this research is to create an AI-based video interview agent capable of predicting communication skills and personality traits by leveraging a multimodal analysis framework. This framework processes three key data types: textual, audio, and video, allowing for a comprehensive evaluation of candidate performance. The architecture integrates machine learning models for each data type, ensuring a holistic understanding of both verbal and non-verbal behaviors. Textual data, derived from candidate responses, is processed through natural language processing (NLP) techniques, while audio data focuses on speech characteristics like tone, pace, and clarity. Video data analyzes non-verbal cues such as facial expressions, body language, and gestures.

The data from these streams is processed simultaneously, and insights are combined into a unified assessment. Each data stream is assigned to a specialized machine learning model designed for that modality, which ensures precise interpretation of human behaviors. To achieve this, interview data is collected from various candidates across different job roles, ensuring diversity and accuracy in model training. Video, audio, and text

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:11/November-2024 Impact Factor- 8.187 www.irjmets.com

transcriptions form the core of the data, which undergoes preprocessing steps to ensure quality and consistency. For instance, tokenization and noise-reduction techniques are applied to text and audio data, while facial landmark extraction is used for video analysis. Once preprocessed, the data is fed into machine learning models trained on domain-specific datasets. The results are aggregated through an ensemble learning system that balances the importance of each modality based on the specific job role being evaluated.



III. MODELING AND ANALYSIS

1. Resume Analysis

The Resume Analysis module will primarily rely on BERT for processing and evaluating candidate resumes. BERT's ability to understand complex sentence structures and extract meaningful context enables it to analyze resumes and identify critical details, such as skills, work experience, and education. By comparing the resume content against job descriptions, BERT can evaluate the alignment between a candidate's qualifications and the requirements of a specific role. This goes beyond keyword matching, as BERT's contextual understanding helps assess the relevance of candidate experiences, even if worded differently. The result is a detailed report, highlighting areas where the candidate's resume excels or needs improvement.

2. Interview Processing

In the Interview Processing module, a combination of BERT, Wav2Vec, and OpenCV will work together to provide a holistic analysis of the candidate's performance during the mock interview. Wav2Vec will convert the candidate's spoken responses into text, which will then be analyzed using BERT to assess the quality, clarity, and relevance of their answers. BERT's fine-tuning capabilities allow it to evaluate whether the candidate's responses are well-structured and appropriate to the context of the questions asked.

Simultaneously, OpenCV will analyze non-verbal communication. OpenCV, which is specialized in facial recognition and emotion detection, will monitor the candidate's facial expressions to gauge their emotional state and confidence levels during the interview. OpenCV, known for its powerful real-time video processing capabilities, will focus on analyzing the candidate's body language, gestures, and posture. It will track subtle movements such as hand gestures, nodding, or signs of nervousness, providing deeper insight into how engaged and composed the candidate appears. This combination of verbal and non-verbal analysis will offer a rich evaluation of the candidate's communication skills and overall demeanor.

3. Feedback Analysis

The Feedback Analysis module will aggregate all the insights from the interview and provide detailed feedback to the candidate. BERT will summarize the content of the responses, offering constructive criticism on the clarity, coherence, and relevance of the answers. It will identify strengths in how the candidate addressed the

questions, as well as areas for improvement, such as providing more concise answers or elaborating on key points.

Wav2Vec will assess vocal performance, offering feedback on aspects such as tone, pitch, and speech clarity. It will help identify whether the candidate spoke with confidence and maintained a steady, clear voice throughout the interview. On the non-verbal front, OpenCV will provide insights into the candidate's facial expressions and body language. This includes evaluating whether the candidate maintained appropriate eye contact, exhibited positive body language, and appeared engaged or nervous. These models work together to generate a well-rounded, data-driven analysis that candidates can use to refine both their verbal and non-verbal communication for future interviews.

4. Algorithm

The study utilizes several advanced machine learning models to capture verbal, audio, and visual aspects of a candidate's interview performance. The text responses of candidates are analyzed using BERT (Bidirectional Encoder Representations from Transformers). BERT, known for its bidirectional context understanding, captures the full meaning of words by analyzing them in both directions within a sentence. This makes it ideal for evaluating complex responses in job interviews. BERT's ability to fine-tune on domain-specific tasks adds an extra layer of precision, ensuring it is adapted to the nuances of job-related answers. BERT is chosen over alternatives such as ERNIE and XLNet due to its more mature ecosystem and easier implementation, which are crucial in real-world applications like interview analysis.

For audio analysis, Wav2Vec, developed by Facebook AI, is employed to convert speech to text, enabling the system to analyze vocal features like tone, pitch, and clarity. Wav2Vec's self-supervised learning model excels in environments with limited labeled data, making it an efficient choice for handling raw audio. Its end-to-end architecture allows direct audio-to-text conversion, reducing the need for intermediate processing. The model's noise robustness further ensures that it can handle interview data recorded in less-than-ideal conditions. Wav2Vec was selected over models like Espresso and DeepSpeech for its superior ability to learn from raw data and its transformer-based architecture, which captures long-range dependencies more effectively.

To assess facial expressions and non-verbal cues, DeepFace is used, leveraging convolutional neural networks (CNNs) for high-accuracy facial recognition and emotion detection. DeepFace is known for its near-human-level accuracy, making it suitable for assessing emotions, confidence, and engagement during interviews. Its robustness in handling variations in lighting and facial angles ensures reliable performance, even in diverse interview settings. DeepFace is preferred over models like FaceNet and Azure Face API due to its ability to handle real-world variations and its flexibility in deployment, offering deeper customization and data privacy controls.

OpenCV, a widely used computer vision library, handles body language and gesture analysis. With its extensive functionality in real-time video analysis, OpenCV enables the system to capture subtle movements and gestures that may provide additional insight into a candidate's engagement and demeanor. OpenCV's cross-platform support and real-time processing capabilities make it an ideal choice for this task, outperforming alternatives like TensorFlow and Scikit-Image, which are either more general-purpose or less optimized for real-time video processing.

IV. CONCLUSION

This research demonstrates the potential of ML models, particularly OpenCV and related technologies, in providing objective analysis of job interview metrics. While these tools show promise in enhancing the hiring process, they should be viewed as supplements to, rather than replacements for, human judgment. Future research should focus on refining these models, addressing ethical concerns, and developing practical implementations that can improve hiring outcomes while maintaining fairness and transparency.

V. REFERENCES

[1] Conrad D, Newberry R (2011) 24 business communication skills: attitudes of human resource managers versus busi ness educators. Am Commun J 13:4–23

[2] Hough LM, Dilchert S (2017) Personality: its measurement and validity employee selection. In: Farr JL, Tippins NT (eds) Handbook of employee selection. Routledge, New York, pp 298–325

[3] Spitzberg BH, Adams TW (2007) CSRS, the conversational skills rating scale: an instructional assessment of interper sonal competence. National Communication Association, Washington, DC

[4] Mohammadi G, Vinciarelli A (2012) Towards a technology of nonverbal communication: vocal behavior in social and afective phenomena. In: Gokcay D, Yildirim G (eds) Afective computing and interaction: psychological, cognitive and neuroscientifc perspectives. IGI, Pennsylvania, pp 133–156

[5] McLarnon MJW, DeLongchamp AC, Schneider TJ (2019) Faking it! Individual diferences in types and degrees of fak ing behavior. Pers Individ Difer 138:88–95

[6] Bambacas M, Patrickson M (2009) Assessment of communication skills in manager selection: some evidence from Australia. J Manag Dev 28:109–120

[7] DeGroot T, Gooty J (2009) Can nonverbal cues be used to make meaningful personality attributions in employment interviews? J Bus Psychol 24:179–192

[8] Nikolaou I, Foti K (2018) Personnel selection and personality. In: Zeigler-Hill V, Shackelford T (eds) The SAGE hand book of personality and individual diferences. SAGE, London, pp 659–677

[9] Torres EN, Mejia C (2017) Asynchronous video interviews in the hospitality industry: considerations for virtual employee selection. Int J Hosp Manag 61:4–13