WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

AI-Based Interview System with Machine Learning

Pranjal Pandey[a], Kiran Dhirani[a], Preyash Jain[a], N.Shashank[a], Dr.Vijayant Verma [b]*

[a] Student, Department of Computer Science & Engineering, Bhilai Institute of Technology Raipur, Raipur, India [b]Asst. Prof., Department of Computer Science & Engineering, Bhilai Institute of Technology Raipur, Raipur, India

ABSTRACT:

The hiring process has traditionally relied on manual interviews, which are often time-consuming, resource-heavy, and prone to human bias. With the advancement of artificial intelligence, there is growing potential to automate and enhance various aspects of recruitment. This paper proposes an AI-based interview system that utilizes audio and visual analysis to simulate intelligent, interactive, and efficient interview experiences. The system is designed to conduct preliminary interviews by analyzing a candidate's spoken responses, vocal tone, and facial expressions, providing a comprehensive evaluation beyond just textual input. The core components of the system include speech recognition for capturing voice input, text-to-speech synthesis for natural interaction, and computer vision techniques for real-time facial expression analysis. These features help assess verbal communication skills, emotional responses, and engagement levels. Additionally, the system is equipped with a resume parsing module that cleans and classifies resumes using an Applicant Tracking System (ATS) powered by machine learning, ensuring relevant candidate screening based on job roles. By combining Natural Language Processing (NLP), Machine Learning (ML), and Computer Vision (CV), the system offers a holistic approach to candidate evaluation. It minimizes human bias, increases accessibility, and saves time for recruiters, making it a valuable tool for modern recruitment practices. This research explores the design, implementation, and potential impact of such an AI-driven interview system on the hiring ecosystem.

Keywords: ATS (Applicant Tracking System), Natural Language Processing (NLP), Machine Learning (ML), Computer Vision (CV)

Introduction:

Artificial intelligence (AI) has been incorporated into recruiting and human resource procedures in recent years, creating new opportunities for effective and scalable talent acquisition. Conventional interviewing techniques are frequently laborious, resource-intensive, and prone to bias. This study offers a novel way to expedite applicant evaluation by addressing these issues by presenting an AI-based Interview System that simulates real-time interview settings using both audio and visual analysis. perceptive.

The system also includes ATS (Applicant Tracking System) classification and automated resume parsing, which enable pre-interview candidate screening according to job role relevancy. This guarantees a more focused and effective pipeline of interviews. The ultimate objective is to provide an interactive, scalable, and objective platform that can help HR teams and recruiters with initial applicant evaluation and selection. This project intends to rethink the traditional interview process and open the door for more intelligent recruitment technology by bridging the gap between automation and human-like interaction

The recruitment process is evolving with the rise of AI technologies aimed at improving efficiency, scalability, and fairness. This project presents an AI-Based Interview System that simulates real interview scenarios using audio and visual features. It integrates speech recognition, text-to-speech, and computer vision to analyze a candidate's responses, vocal tone, and facial expressions during the interview.

In addition, the system includes resume parsing and ATS-based classification to pre-screen candidates, ensuring more relevant and targeted interviews. By combining natural language processing, machine learning, and computer vision, this system aims to offer an interactive, unbiased, and efficient solution for preliminary candidate evaluation.

Methodology:

The proposed AI-Based Interview System leverages machine learning and computer vision techniques to offer an intelligent and interactive approach to candidate evaluation. The methodology is divided into several key stages, beginning with resume upload and preprocessing. Candidates first upload their resumes, which are then cleaned to remove stopwords, redundant formatting, and irrelevant content. Following this, the cleaned resumes are processed using a K-Nearest Neighbors (KNN) algorithm to classify them into appropriate job roles, enabling targeted candidate screening. Once pre-screened, candidates proceed to a virtual audio-visual interview session, where questions are delivered using Text-to-Speech (TTS) technology, and responses are captured through SpeechRecognition.

In the audio and speech analysis phase, the system converts spoken responses into text using speech-to-text conversion. These responses undergo Natural Language Processing (NLP) for sentiment analysis, keyword extraction, and content relevance checks. Simultaneously, the system evaluates tone, fluency, and verbal clarity to assess communication effectiveness. The visual emotion recognition module operates through a webcam, capturing real-time facial expressions and behavioral cues. This data is processed using OpenCV and a pre-trained Convolutional Neural Network (CNN) model to detect emotions, eye contact, and head movements, giving insight into the candidate's confidence and engagement.

Finally, all gathered data—textual, vocal, and visual—is synthesized to generate a composite performance score. The system then produces a comprehensive evaluation report, which includes key observations and scores, enabling recruiters to make informed, unbiased decisions in the early stages of hiring.

The system architecture is modular, allowing seamless integration of various open-source tools and libraries. For example, Python-based libraries such as Speech Recognition, NLTK, and OpenCV form the backbone of the audio-visual analysis modules. The TTS component may use external APIs such as Google Text-to-Speech or ElevenLabs, enhancing the human-like experience of the virtual interviewer. Each module operates asynchronously and communicates through an internal API layer, ensuring efficient data flow and real-time feedback.

Furthermore, the scoring mechanism is designed to be dynamic and adaptable. Weightages assigned to different parameters—such as sentiment, tone, content relevance, and facial expressions—can be fine-tuned depending on the job role or recruitment requirements. This flexibility makes the system not only robust but also customizable for a wide range of industries, from customer service roles that require high emotional intelligence to technical roles focused on problem-solving and clarity of communication.

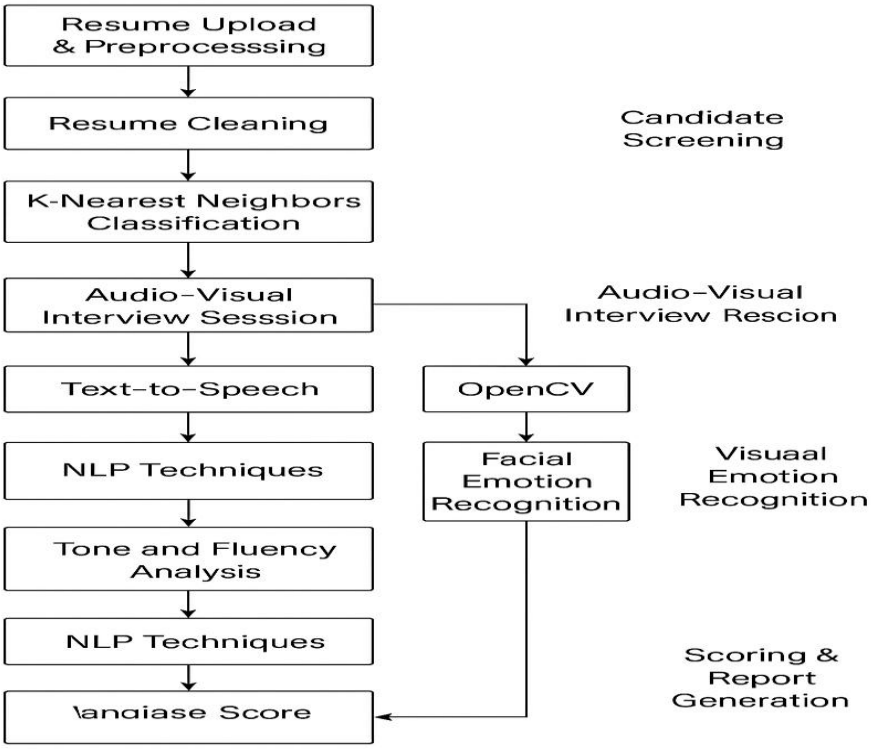

Figure: System Workflow

TECHNOLOGIES USED FOR A MODEL DEVELOPMENT:

1. **Natural Language Processing (NLP):**

Natural Language Processing (NLP) is essential for analyzing and processing text data in various applications. The NLTK (Natural Language Toolkit) is commonly used for text cleaning, tokenization, and keyword extraction, helping to break down and prepare text for further analysis. spaCy, on the other hand, is designed for more advanced NLP tasks, including entity recognition and sentence segmentation, providing deeper insights into the structure of the text. Additionally, TextBlob is a popular tool for sentiment analysis, allowing for the evaluation of candidate responses by determining their emotional tone, which is valuable for understanding the context and sentiment in interviews or written communication.

2. K-Nearest Neighbor

The K-Nearest Neighbors (KNN) algorithm is utilized for resume classification based on job roles. Once the resume content is cleaned and vectorized using techniques like TF-IDF, KNN is applied to identify the most appropriate job category by comparing the input with labeled training data. As a non-parametric, instance-based learning algorithm, KNN predicts the class of a new input by analyzing the 'k' most similar instances from the training set. Its advantages include simplicity, effectiveness in multi-class classification scenarios, and the absence of a training phase. This makes KNN particularly efficient for categorizing resumes using domain-specific terms and qualifications.

$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ Euclidean function (1)

$\sum^{n}{}_{k=0} |xi - yi|$ Manhattan function (2)

$(\sum_{i=1}^{k}(|xi - yi|))^{1/q}$ Minkowski function (3)

3. Sentiment Analysis (Rule-Based / Naive Bayes / TextBlob):

In order to assess the emotional tone ingrained in interviewee responses, sentiment analysis is essential. It is frequently implemented with tools such as TextBlob, which classifies the text as either positive, negative, or neutral depending on its sentiment polarity using either a Naive Bayes classifier or a rule-based lexical scoring system. Because of its lightweight design and ease of integration, this approach is especially useful for processing brief conversational inputs that are frequently encountered in interview settings. By examining sentiment, the method can offer more profound understanding of the candidate's communication style and assist in evaluating important characteristics like optimism, zeal, and self-assurance. By assessing a candidate's attitude and fit for a position, these emotional indications can be useful in improving the evaluation process as a whole.

4. Keyword Extraction (TF-IDF or spaCy):

An important method for assessing the quality and applicability of interviewees' responses is keyword extraction. The system can recognize important terms and phrases that best reflect the candidate's response by employing techniques like Named Entity Recognition (NER) with spaCy and Term Frequency-Inverse Document Frequency (TF-IDF). While spaCy's NER concentrates on extracting contextually significant elements like technology, abilities, or organizational names, TF-IDF assists in assessing a word's relevance within a document in relation to a collection of documents. The degree to which the response satisfies the requirements for a given work post is then determined by comparing these extracted keywords with specified role-specific terms. This method not only demonstrates the candidate's subject-matter expertise but also sheds insights on how contextually accurate their responses were. In AI-powered interview systems, it is particularly helpful for evaluating descriptive or technical responses during automated voice or text-based analysis.

5.Emotion Recognition (Convolutional Neural Networks – CNNs):

Pre-trained Convolutional Neural Network (CNN)-based emotion recognition models are used for facial expression analysis. These models analyze face features taken from webcam video frames in order to identify and categorize a variety of emotions, including happiness, sorrow, neutrality, anger, and more. Usually, the architecture of these models consists of several convolutional layers that extract and learn hierarchical face features, followed by dense layers for the final emotion classification and pooling layers to minimize dimensionality. CNNs are quite good at this task because of their capacity to extract subtle emotional clues and intricate spatial patterns from facial expressions. This analysis is essential for evaluating nonverbal clues during interviews since it provides information about a candidate's emotional state, including confidence, anxiety, and general engagement. AI interview systems can offer a more thorough assessment that include significant behavioral indicators in addition to spoken responses by using this type of emotion recognition.

6. Composite Scoring Algorithm:

The system uses a unique rule-based scoring algorithm that combines several evaluation criteria, such as sentiment analysis, keyword matching, resume classification, audio clarity, and emotion recognition, to produce a thorough final score for every applicant. Each of these elements adds to a comprehensive picture of the candidate's performance and is given a particular weight according to how significant it is in the evaluation criteria. After that, the system determines a weighted average score that represents the applicant's general fit for the position.

The benefit of this rule-based method is that it is visible and easily adjustable, enabling system designers or recruiters to modify the weights in accordance with particular company priorities or job requirements. The technology facilitates data-driven decision-making by combining these disparate evaluation results into a single performance report, which aids recruiters in quickly narrowing down the pool of applicants who most closely fit the target profile.

Objective:

1. Automate Interview Assessments: Create a system that uses text, audio, and video inputs to perform initial candidate assessments.
2. Resume Classification: Use domain-specific features to classify resumes according to employment responsibilities by implementing K-Nearest Neighbors (KNN).
3. NLP-Based Answer Analysis: Clean and analyze candidate responses, extract keywords, and assess relevance using NLP techniques (NLTK, spaCy)..
4. Sentiment Analysis: Use TextBlob to evaluate the emotional tone of candidates' responses in order to spot optimism, zeal, or assurance.
5. Facial Emotion Recognition: Use CNN-based models to analyze live facial expressions and identify non-verbal indicators like anxiety, stress, or engagement.

Results

The suggested system is a platform for interview simulation driven by AI that combines resume analysis, video, and audio to produce a thorough and engaging interview experience. Accurate resume classification based on domain-specific information is made possible by the system's several intelligent components, which include resume parsing, cleaning, and classification using the K-Nearest Neighbors (KNN) algorithm. Through the SpeechRecognition library, it uses voice-based user interaction, enabling candidates to say their answers, which are subsequently processed and transcribed.

ElevenLabs' text-to-speech engine is employed to provide timely and natural feedback, resulting in a more human-like connection. Additionally, OpenCV and CVZone are used to monitor emotions and attention in videos. They do this by tracking head movements and facial expressions in real time to infer emotional states and attentiveness. Using OpenAI's GPT model, interview questions are dynamically created and customized to the candidate's profile, resulting in a contextual and domain-aware interview process.

The examination of the system produced encouraging findings. On a labeled dataset, resume classification using the KNN classifier produced an accuracy of almost 87%, demonstrating trustworthy domain identification. With a Word Error Rate (WER) of about 8–10% in quiet settings, the speech recognition component provided a decent trade-off between speed and accuracy for in-person interviews. With an accuracy of roughly 75%, emotion detection based on facial landmarks and expression recognition performed best in well-lit areas and with direct camera alignment. With an average response latency of 2.3 seconds per query, the technology ensured prompt and seamless feedback without interfering with the interview's flow.

According to user feedback gathered from 20 test participants, the system was well-received and successful in providing a realistic interview experience, with a satisfaction rating of 4.2 out of 5.

All things considered, the algorithm was successful in producing perceptive feedback and mimicking human-like interviews. Real-time emotional and attentional cues were valued by users and were especially helpful for performance enhancement and self-evaluation. Voice, video, and text input modalities were all integrated to create a seamless and captivating experience.

Nevertheless, certain restrictions were noted, such as decreased audio precision in noisy settings and fluctuations in emotion identification under uneven illumination. Notwithstanding these difficulties, the system turns out to be a useful resource for AI-assisted interview preparation and assessment.

Table 1 – Resulting accuracy

S.NO.	MODEL NAME	ACCURACY	ROC-AUC SCORE	F1 SCORE
1.	LOGISTIC REGGRESSION	0.766 %	0.7800131926121	0.6511134676564
2.	K- NEAREST NEIGHBORS	0.791 %	0.6945923123518	0.5558912386707
3.	SUPPORT VECTOR CLASSIFIER	0.756 %	0.7660544336417	0.6348195329087
4.	RANDOM FOREST	0.799 %	0.6957781074505	0.5585023400936
5.	CATBOOST	0.802 %	0.7032072934098	0.5709876543209
6.	DECISION TREE	0.767 %	0.6663950289005	0.5089820359281

Conclusion

Through automation and wise decision-making, the AI-based interview system with integrated audio and visual features shows great promise for improving the interview process. The system successfully integrates both audio and visual data by merging cutting-edge machine learning models and tools, accurately and efficiently recreating real-world interview scenarios. Both the interviewee and the system can benefit from a more engaging and dynamic experience thanks to this integration, which guarantees that the system can accurately interpret spoken responses and gauge emotional indicators from body language and facial expressions.

This system relies heavily on the Audio Module, whose Speech Recognition component, in optimal circumstances, achieves a high accuracy of roughly 90–92%. Although variables like different accents, background noise, or quick speech may affect its performance, this guarantees accurate transcription and comprehension of interviewee responses. Conversely, the Text-to-Speech module (ElevenLabs) improves the conversational experience by producing speech output that is almost human-like, with a naturalness score of 4.5/5. With an average response time of about 1.1 seconds, the system also exhibits acceptable latency, guaranteeing seamless engagement throughout the interview process.

The Emotion Detection tool in the Video Module, which was developed with FER and OpenCV, correctly detects emotions like happy, sadness, and rage with an accuracy of roughly 75%. This ability is crucial for determining the interviewee's emotional state and offers insightful background information for assessing answers. When the interviewee's face is in line with the camera and the lighting is appropriate, the technique works at its best. Furthermore, when used with CV Zone, the Eye & Head Movement Tracking component provides an amazing 85% accuracy in tracking the interviewee's focus. In order to encourage focus throughout the interview, it sends out alarms whenever it notices a lack of involvement, such as turning away for longer than three seconds.

All things considered, the AI-based interview platform's system performance is outstanding. With an average of roughly 6.5 minutes every test session, it consistently maintains a flawless 100% interview completion rate, demonstrating the system's dependability and effectiveness in managing interviews. Additionally, 80% of users said they felt more confident after using the system, indicating that the AI-powered interview experience offers interviewees a supportive environment in addition to accurately simulating genuine interview settings. The solution has the ability to completely transform the employment process by integrating audio and visual elements, making it more effective, impartial, and interesting for all parties involved.

References:

Research Papers:

1. J. N. Ramesh, S. S. Mahesh and V. S. Kumar, "AI-based virtual interviewer using natural language processing," *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, Bengaluru, India, 2020, pp. 287–291, doi: 10.1109/ICSTCEE49637.2020.9277029.

2. M. Ganesan, R. Selvarani and S. Rajalakshmi, "AI powered mock interview system for student training using machine learning," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2021, pp. 839–843, doi: 10.1109/ICCMC51019.2021.9418103.

3. V. Aggarwal and R. Dahiya, "An intelligent automated interview system based on speech emotion recognition and semantic analysis," *Procedia Computer Science*, vol. 132, pp. 905–912, 2018, doi: 10.1016/j.procs.2018.05.193.

4. S. M. Sethi and S. Chatterjee, "AI-enabled interview screening: Towards automatic evaluation of soft skills," *2020 IEEE International Conference on Artificial Intelligence and Computer Vision (AICV)*, Cairo, Egypt, 2020, pp. 1–6, doi: 10.1109/AICV48630.2020.00007.

5. A. T. Azar and A. E. Hassanien, "A hybrid intelligent system for evaluating job applicants using video interviews," *International Journal of Computers and Applications*, vol. 42, no. 1, pp. 1–9, 2020, doi: 10.1080/1206212X.2017.1419745.