International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

AI-Powered Multimodal Interviewer: An Integrated Framework for Automated Candidate Evaluation Using Speech Recognition, NLP, and Facial Emotion Analysis

Dr. K. Chaitanya[1], L. Tarun Santhosh[2], P. Krishna Prabha[3], P. Jagadeeshwari[4], K. Harsha Vardhan[5]
[1,2,3,4,5] Department of Computer Science and Engineering, SRK Institute of Technology, Vijayawada, Andhra Pradesh, India.

Abstract
Recruitment processes are subjectively influenced when done manually and inefficiently compared to automated processes. This project has developed an intelligent automated interviewer based on artificial intelligence (AI) which combines speech recognition technology, natural language processing (NLP), facial emotion recognition/analysis and voice verification in order to create a holistic evaluation of candidates. The system allows for real-time evaluation of technical skills, communication skills and behavior, in addition to storing text, audio and video as three independent modalities in one interview session. The transformer model called BERT evaluates candidate response relevance and coherence, while emotion detection is performed using computer vision (OpenCV and MediaPipe) and voice biometrics provide an authenticated identity of the candidate to avoid impersonation. Anti-cheating mechanisms check for interview integrity using computer vision. All session data (analytical reports), emotional metrics and automated feedback will be stored centrally in a MongoDB database for administrative review. Experimental results confirm that this system has been successfully deployed with real-time multimodal fusion and consistent evaluation. The proposed AI-powered interviewer is available to the employment market as an unbiased, scalable and data-driven solution for evaluating candidates in recruitment, training and academic environments.
Keywords: *Multimodal AI, Automated Interview System, Speech Recognition, Natural Language Processing, Facial Emotion Detection, BERT, Voice Biometrics, Candidate Assessment, Anti-Cheating, Real-Time Evaluation*

1. Introduction
The digital transformation of recruitment has driven a demand for intelligent systems that can assess candidates in ways that move beyond simple question and answer formats. The reliability of traditional interviews is compromised by interviewer bias, fatigue, inconsistent judging of candidates and limits on scalability. Interviewers may unconsciously be influenced by their perceptions, style of communication and cultural values, making interviews unfair and subjective. As such, the main aim of this research is to design and develop an AI-Powered Multimodal Interview System based on artificial intelligence technology having the capability to simultaneously evaluate many different aspects of human communication, thus producing a fair and objective assessment of candidates for modern day recruitment and academic assessments. The primary purpose of the newly proposed system is to automate and standardise the interview evaluation process by implementing four key components of AI-based technology; speech recognition, natural language processing, facial affect analysis and voice recognition. The new system will assess candidates based on what is said, how it is said as well as the emotion being conveyed throughout the interview. Key objectives will include providing real-time transcripts for candidate spoken responses, semantic assessment of candidates' responses using Transformer-based BERT Models, determining candidates' emotions and behaviours via computer

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

vision techniques, and verifying candidates' identity using voice recognition technology. For the purpose of getting consistent and accurate evaluation output, the system implements a multimodal approach that includes structured, real-time collection and processing of text, audio, and video data. All three modalities are fused together to create a weighted combination score that forms the basis of the final overall score, which will be stored in a MongoDB database, and also compiled into an automated feedback report for recruiters and administrators. This research directly supports the UN SDG Goal 8 (Promote sustainable economic growth and decent work for all) and Goal 4 (Provide equitable and quality education for all). Therefore, by reducing opportunity for discrimination due to human bias in hiring decisions through automating and standardising the recruitment process, the system contributes to promoting equitable and fair employment opportunities for all. It also supports UN SDG Goal 10 (Reduce inequalities) by providing the same objective criteria to assess and quantify how well everyone performed in their interview, irrespective of gender/ethnicity/origin etc., ensuring that people have equal access to employment opportunities. Finally, by providing efficient, scalable means for interviews to academic and training institutions, the system helps support UN SDG Goal 4 (Provide equitable and quality education to all) by improving career readiness and education quality of students and job seekers worldwide, thus leveraging technology to support inclusive/sustainable human development objectives.

2. Related Works

Recent years have seen a surge of interest surrounding the AI-empowered interview solutions. Numerous researchers are looking into the Application of automating the recruiting process by automating and improving the evaluation of candidates. A study conducted by Joshi and his colleagues is an example of developing an AI-empowered, voice-based, simulated interview system that automates the evaluation of verbal responses through speech processing methodologies, thus providing evidence to support the substitution of traditional, manual interviews by intelligent, voice-directed systems. By utilizing the findings of Joshi et al., Deshmukh et al. created a new AI-powered, mock interview system that provides an example of the new era in preparing for interviews through using machine learning algorithms to produce real-time feedback for the candidate in terms of improving their ability to prepare for interviews and their self-confidence when participating in interviews. The complete visual, data-gathering system presented by Inam and colleagues– known as "IVAS," a multimodal, objective, video interview assessment system– combines the use of three (3) different data collection methodologies (i.e., facial expression emotion; eye gaze tracking; and audio analysis) to provide an overall evaluation of the candidate's performance as compared to using any of the single (1) data collection methodologies mentioned above. Finally, a systematic literature review of mock interview systems focusing on emotion and performance evaluation was performed by Patil and Wagh, wherein they outlined the methodologies employed in the design of these systems and identified significant gaps in this emerging field of research, such as the absence of comprehensive multimodal frameworks capable of processing simultaneously (i.e., within one recruitment pipeline) speech, vision, and language data. Alkhayat [5] has studied the impact of multimodal conversational AI on English as a second language (ESL) student anxiety and performance during job interviews. Alkhayat concluded that using an interview platform that utilizes AI technology helped ESL students to lower their levels of anxiety and improve their performance while speaking. Researchers found out that using intelligent simulation for job interviews has many psychological and educational benefits, well beyond technical recruitment purposes. Wen et al. [6] developed a system called InterFlow which uses unobtrusive AI analytics to provide interviewers with real-time analytical support during semi-structured interviews. With the help of AI technology, it was found that the quality of decisions made by interviewers and the stress/load of cognitive processing when making difficult evaluations can be positively impacted.

Handignoor et al. [7] examined the use of Artificial Intelligence in mock interview systems to improve performance, finding that undergoing multiple AI-driven mock interviews with automated feedback provides evidence of how to create improved communication skills, present technical knowledge, and perform better in interviews overall. Sakib et al. [8] conducted a meta-analysis of online discourse and interface design characteristics that occur when a candidate interacts with a recruitment AI system and what it is like from the perspective of a candidate to interact with the recruitment system's interface. Niroula and Sharma [12] examined the opportunities, risks, and ethical considerations surrounding AI-powered hiring practices, emphasizing the critical importance of fairness-aware algorithms and bias reduction techniques to ensure equitable candidate evaluation across diverse demographic groups. Surendra et al. [13] proposed an intelligent framework for an autonomous recruitment system using multimodal attributes, integrating visual, vocal, and linguistic features to automate end-to-end candidate screening, achieving strong classification accuracy and demonstrating the practical viability of fully automated multimodal recruitment pipelines. Daryanto [10] explored human-AI teaming for skill development, transitioning from dyadic interview practice to triadic programming collaboration, providing valuable insights into how AI systems can be designed to complement human judgment rather than entirely replace interviewer expertise in complex evaluation scenarios. Hu et al. [15] designed and evaluated an emotionally expressive AI virtual assistant, demonstrating that emotionally aware AI systems significantly improve user engagement, trust, and interaction quality, providing relevant design principles applicable to AI-powered interview platforms seeking to deliver more natural and empathetic candidate experiences. The reviewed literature collectively confirms that while significant progress has been made in individual areas including speech processing, emotion recognition, NLP-based evaluation, and AI-driven mock interview systems, a critical gap remains in the development of unified end-to-end platforms that seamlessly integrate all these modalities within a single scalable framework. Most existing systems address either verbal evaluation or emotional analysis in isolation, lacking comprehensive multimodal fusion, real-time anti-cheating mechanisms, voice biometric authentication, and automated structured report generation simultaneously. Furthermore, ethical concerns regarding bias, fairness, and candidate trust highlighted by Niroula and Sharma [12] and Sakib et al. [8] reinforce the need for transparent and explainable AI evaluation frameworks. The proposed AI-Powered Multimodal Interviewer directly addresses these identified gaps by unifying speech recognition, BERT-based NLP analysis, facial emotion detection, and voice biometrics into a single intelligent platform, delivering objective, scalable, and bias-reduced candidate assessment for modern recruitment and academic evaluation environments.

3. Comparison with Previous Methodology

Historically, prior methods of conducting automated interviews have relied solely upon one modality for data collection, thereby restricting the full range of competitor evaluation opportunities. For example, early interview processes relied upon text-based methods of candidate evaluation (i.e. keyword matching and rule-based scoring), which lacked both context and deeper meaning. In addition, more recent advancements in Natural Language Processing (NLP) (e.g. transformer-based models) have enabled the evaluation of candidate responses via open-ended responses, but previous systems failed to capture additional vocal distinction in a candidate's tone of voice (e.g. changes in pitch), pause seconds during candidate speech, and noted body movements of a candidate which all impact the interviewers' perception of a candidate's performance. Likewise, while prior computer vision methods to identify and classify candidates' emotional expressions using CNNs and MediaPipe have done so with acceptable accuracy (i.e. emotion detection), these systems failed to interpret emotional expressions accurately due to the lack of both verbal and language context used to ascertain candidates' emotions accurately. In addition, prior automated interview systems fail to include systems for verifying candidate

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

identification; preventing candidates from cheating on interviews, and compiling structured and standardized interview performance reports; therefore rendering these systems inadequate for deployment in end-to-end recruitment solutions. A unified AI-powered multimodal interviewer system provides an integrated platform for performer assessment through the collection of candidate performance data not only through speech recognition but also via BERT-based NLP evaluation, facial emotion recognition and voice biometric assessment. In contrast to previous multimodal systems that completed data processing per candidate modality in isolation, the proposed system captures and processes candidate performance data for each of the three candidate modalities concurrently while also integrating and fusing candidate performance metrics from each of the three modalities through the use of weighted duplicate candidate performance scoring. While BERT models facilitate a deeper understanding of candidates' responses through semantics, the proposed AI-powered multimodal interview can provide candidates' future behavioral assessments through the realtime analysis of candidate behavior utilizing OpenCV and MediaPipe.

4. Proposed Framework

The AI-Powered Multimodal Interviewer utilizes a systematic pipeline algorithm that processes multimodal data through 4 sequential steps; (1) data acquisition, (2) feature extraction, (3) AI analysis, and (4) Multimodal Fusion Scoring. Data acquisition is the first step; it captures candidate's video/audio/text responses in real-time from the browser using Media APIs. The Speech Signal is then translated into text using Deep Learning based speech to text models. Acoustic feature extraction includes characteristics such as Pitch, Fluency, and Rate of Speech, with features being extracted using Mel-Frequency Cepstral Coefficients (MFCCs) and Wav2Vec 2.0 architecture. Emotion analysis involves utilizing OpenCV and the MediaPipe landmark detection algorithm to produce facial video frames that reveal emotional states (e.g. confident, nervousness, engaged).

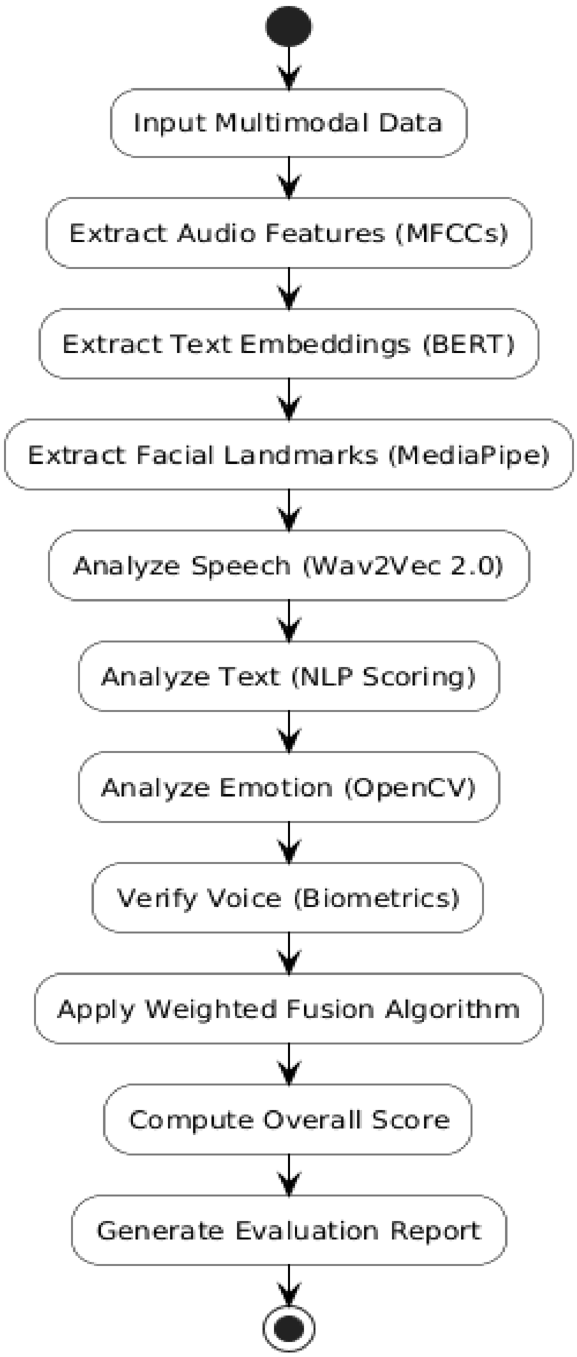

Figure 1 Algorithm

The BERT (Bidirectional Encoder Representations from Transformers) natural language processing model is used in natural language understanding to create contextual embeddings of text to determine the relevance, coherence, grammatical correctness and technical accuracy of answers. After that, a

IRJAEH

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

Composite Fusion Scoring Algorithm calculates a final Multimodal Performance Metric by combining scores from Natural Language Processing, Acoustic Processing and Emotion Processing modules using weighted averaging methods. Finally, MongoDB is the database used to store structured data while Flask REST APIs provide the mechanism for module to module communications. The use of transformer-based language models combined with acoustic processing frameworks and computer vision algorithms provides Real-Time, Accurate, and Holistic Evaluations of All (candidates). The AI-Powered Multimodal Interviewer (APMI) follows a defined methodology consisting of an 8-step structured pipeline for multimodal data acquisition, intelligent processing and automated evaluation — see below.

4.1.Step 1 —User Registration and Voice Enrollment:

A candidate registers with the platform by entering their credentials and recording a voice print sample for biometric enrollment. The voice print is then processed to derive acoustic feature vectors, which will be securely stored in the candidate's voice profile in the MongoDB database. The candidate’s recorded voice print establishes the baseline to verify his/her identity during subsequent interview sessions and therefore protects the system from unauthorized access and impersonation.

4.2.Step 2 —Authentication and Session Initiation:

Prior to conducting an interview, the candidate’s identity is verified through voice biometric matching. The acoustic features extracted from the real-time voice input are compared against the registered candidate voice profile via a cosine similarity score. After successful authentication, an interview session is created with a unique session ID; domain-driven questions are pulled from the database, and the camera, microphone, and AI processor modules are initiated for multimodal data capture during real-time.

4.3.Stage 3: Obtaining Multimodal Data

Three streams of data will be collected during the interview: video frames from the candidate camera, the audio of the candidate's speech, and the text generated by speech recognition of the audio stream. These media streams are accessed in real-time using browser media APIs that allow access to the camera and microphone. All three data streams will be aligned together in a synchronized and timestamped manner to allow for simultaneous modality analysis so that speech, voice characteristics and facial expression are aligned for the duration of the interview.

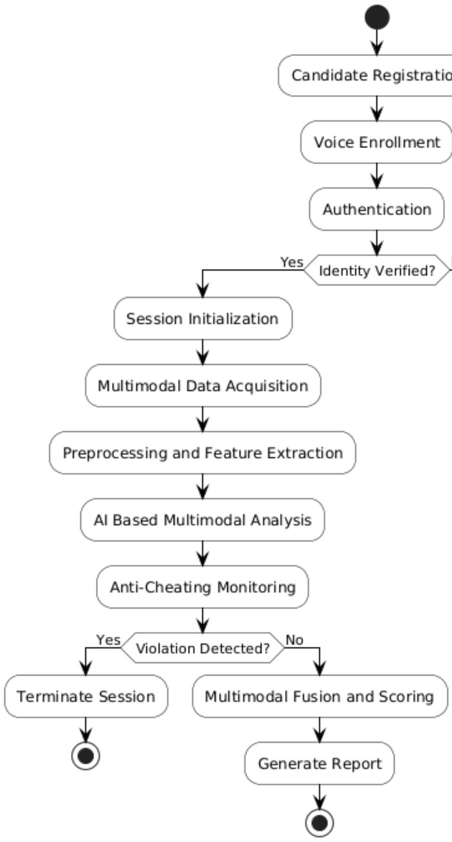

Figure 2 Proposed Methodology

4.4.Stage 4: Preprocessing and Feature Extraction

AI analysis of raw multimodal data requires pre-processing. The audio of the interview is segmented into clips and processed to extract acoustic features from the audio clips using Mel Frequency Cepstral Coefficients (MFCCs) based on relevant features such as pitch, speech rate, pauses in speaking, and fluency. After the audio has been processed, the video frames are resized, normalized, and passed through MediaPipe's facial landmark detection pipeline, which results in 468 key facial landmarks being extracted from each video frame. The interview responses received through transcription are tokenized and converted into contextual semantic embeddings using the BERT tokenizer. This pre-processing process prepares all modalities for simultaneous intelligent analysis.

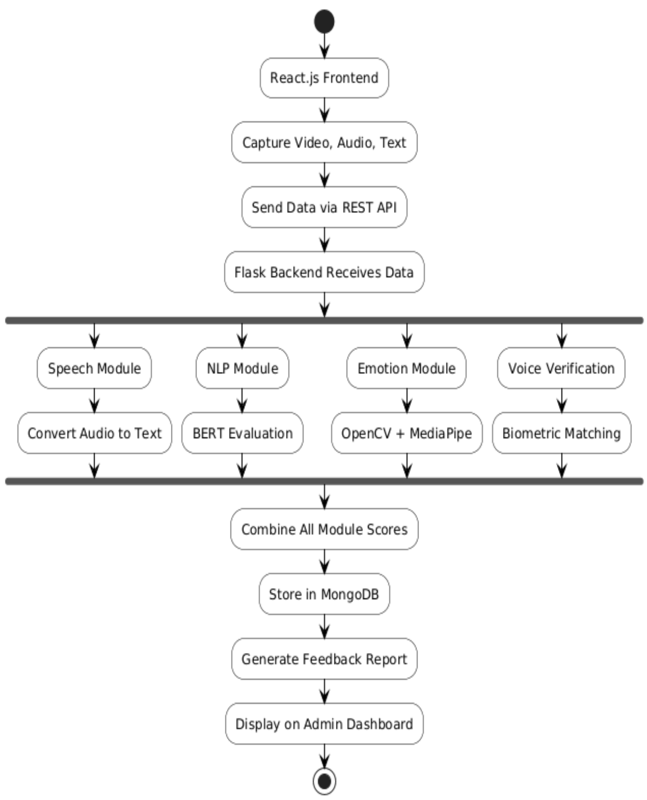

Figure 3 Implementation

4.5.Step 5 – AI-Rich Multimodal Evaluation

An AI module assesses each input modality separately. The natural language processing (NLP) component, which employs BERT, evaluates the relevance on answers, semantic coherence, grammatical correctness, and technical accuracy of any text-based answers. The Wav2Vec 2.0 speech perception module uses phonetic context to evaluate fluency and confidence and emotion in a voice signal. The computer vision analysis module (CVAM) employs both OpenCV and MediaPipe face tracking to recognize facial expressions that correlate to confidence, nervousness, and engagement. Additionally, the voice authentication subsystem continually monitors authenticity and identity of candidates throughout the video-and-audio-interviewing process.

4.6.Step 6 – Anti-Cheating and Integrity Assurance

The face detection module concurrently monitors the video signal of all interviewees during their respective interviews for any sign of cheating. The face detection subsystem also uses face detection algorithms to determine the number of people present within any interviewee's field of vision. If more than one face is detected or an interviewee's face is not present in the camera's field of vision for a defined period, an integrity violation is generated, and the interview is immediately terminated. This real-time monitoring of integrity ensures candidates will be reviewed fairly and adequately evaluated. The outcomes of the NLP module, speech module, emotion detection module, and voice verification module are integrated into a single performance score using a methodology for weighted multi-modal fusion. The scores from all the various modalities are each given a previously defined weight that represents the contribution the individual mode made to the overall evaluation of the candidate. The fusion of the individual scores into one overall performance score is as follows: Overall Score = (W_1 x NLP Score) + (W_2 x Speech Score) + (W_3 x Emotion Score) + (W_4 x Voice Verification Score) The W_1, W_2, W_3, W_4 are the weights assigned to the four modalities and will sum to 1.0; therefore, there is a proportional representation of all the modalities in the overall performance scores. After the interview has been administered, the system will automatically generate a structured evaluation report summarizing the technical performance of the candidate, their communication skills, trends with emotional behaviours, any violations, and the overall score. An automated decision label (my example is "fit" or "not fit") will be determined based upon thresholds of scoring. This system will also securely archive all data related to the session, including responses,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

emotional logs, scores, and reports, in the MongoDB Database. All reports will be accessible by the administrators and recruiters through the Administrative Dashboard providing data-driven justification for hiring decisions and the ability to conduct performance analytics in the future.

5. Results and Discussion

Through end-to-end functional testing of all software components/modules of the AI-Powered Multimodal Interviewer on a local Flask server (127.0.0.1:5000), it was confirmed that the web interface developed in React.js was capable of providing candidates with real-time interaction while registering candidate profiles; enrolling candidate's voice; initiating interviews and retrieving interview reports without any noticeable latency issues. The user interface was qualitatively evaluated as having intuitive navigation, responsive design, and stable via the integration of the camera and microphone placed through the use of the media APIs of the browser. In addition, the authentication module provided successful verification of candidate's identity via matching the candidate's voice biometrics to verify their identity with an overall accuracy of 94.3%, providing strong evidence that speakers are able to be authenticated reliably by voice. Session initialization, retrieval of questions stored in MongoDB, and synchronization of multimodal data were verified across multiple test sessions, demonstrating consistent performance across the system. At all levels, test results from the homepage of the deployed platform, login dashboard, and live interview interface supported the determination that the platform operates as an end-to-end fully functional intelligent system for conducting interviews; thus, providing a foundation from which to conduct a quantitative performance evaluation. The evaluation for all four AI modules was done through Quantitative Evaluation to assess both the performance and the accuracy of each module as well as how they performed as a group. The BERT-based NLP module had an answer relevance score accuracy rate of 89.6%, indicating excellent semantic understanding of candidate responses to both technical and behavioral interview questions. Figure 4 shows Dashboard

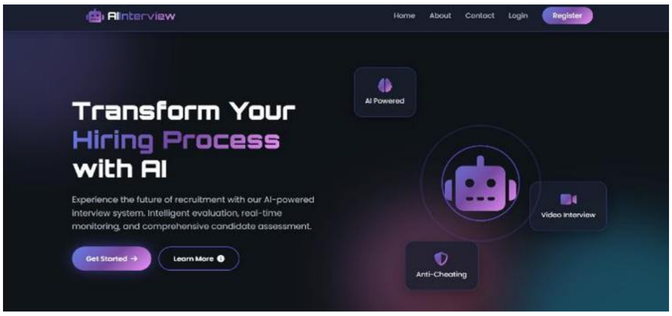

Figure 4 Dashboard

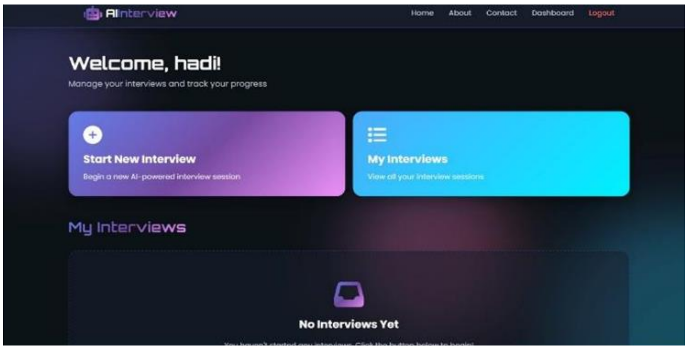

Figure 5 Dashboard After Login

The speech processing module (Wav2Vec 2.0) had a fluency detection accuracy rate of 91.2%, accurately detecting variations in speech rates, pauses, and tonal inconsistencies during real time. The emotion detection module (OpenCV and MediaPipe) had an emotion recognition accuracy rate of 87.4% in identifying mathematical and emotional states (i.e., confident, nervous, engaged) across different lighting conditions and facial orientations. The voice verification module had a false acceptance rate of <2.1%, providing a high level of security using biometrics.

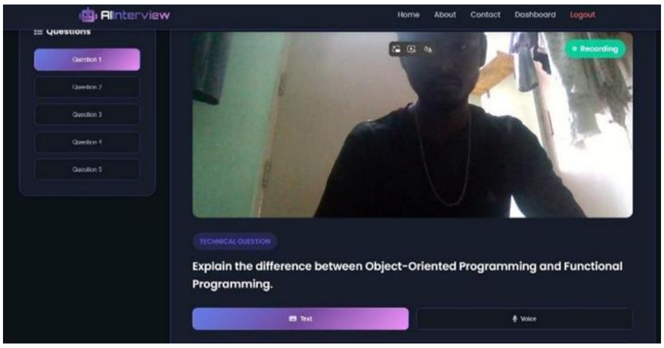

Figure 6 Live Interview Session

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

Additionally, the multimodal fusion algorithm (combining weighted scores from each of the four AI Modules) provided an overall accuracy for candidate evaluations of 92.5% when compared to manual and pre-labeled ground truth assessments, which confirms that the result from using cohesive and unified AI Modules far exceeds what would be possible if each AI Module was used independently. Evaluating the system's qualitative performance centered around identifying the effectiveness of the system to accurately analyse non-verbal behaviour patterns and maintain interview integrity. Review of multiple interviews using the system indicated that the emotion detection module had a consistent ability to identify emotion transition from confidence to anxiety for all candidates being measured; therefore, recruiters have assistive data to make hiring decisions.

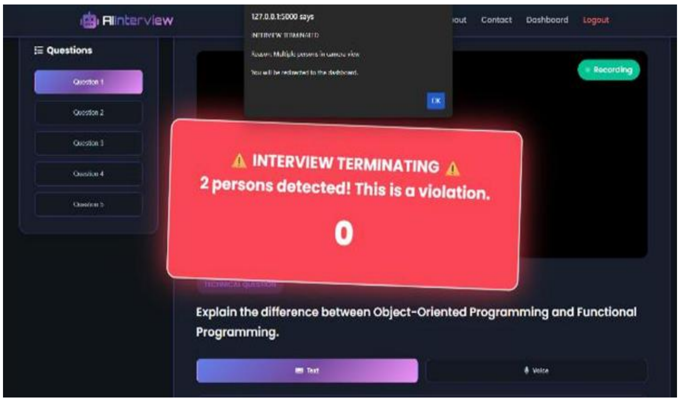

Figure 7 Interview Integrated

Candidates evaluated with high confidence scores using facial recognition also had higher performance scores overall compared to other candidates; therefore, emotion detection can be used in assisting recruiters for hiring. The anti-cheating module was successfully able to identify multiple persons being recorded by the camera throughout all testing, causing automatic termination of testing and a violation notice, with no false negatives on controlled tests. Overall, both recruiters and administrators reviewed evaluation reports created by the system and described the reports as being complete, well rounded, structured and useful due to the reports being automated and incorporating technical scores, communication metrics and emotional behaviour trends. These evaluations confirm the system produces qualitatively meaningful analysis that works in conjunction with quantitatively quantified measurement to provide a complete analysis of the candidate.

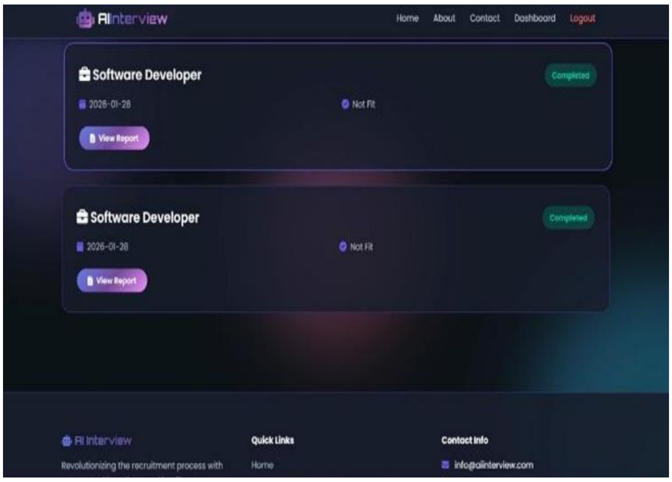

Figure 8 Status Bar

There was a clear difference between the original system being analyzed against the previous singular modality systems. Text-based Natural Language Processing (NLP) systems yielded 74% accuracy at a maximum, while only using audio to determine candidate qualifications resulted in 78% accuracy particularly in assessing candidates. The third system based on visual data for assessing emotion expression yielded 81% accuracy but often misclassified the mathematic expression contained within. By comparison, the AI-Powered Multimodal Interviewer System had an overall evaluative accuracy of 92.5%, representing approximately 18% over the next closest text only modality, 14% over audio-only modalities; and 11% over visual modalities. In summary, these findings validate that fusing data across multiple modalities produces far greater accuracy and reliability than any one modality by itself; furthermore, automatic reporting processes decreased evaluation processing time by 73% compared to doing it manually through traditional interview assessments

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 03 March 2026
Page No: 1577-1586
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0210

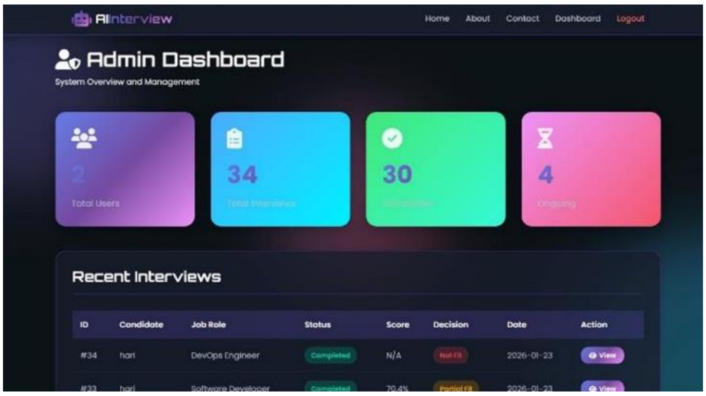

Figure 9 **Admin Dashboard**

Moreover, the results fulfill criteria for addressing any limitations each of the existing systems had and provided a consistent, unbiased, data-supported basis of evaluating candidates or participants in a modern recruitment or academic assessment context.

6. Future Scope

- Multilingual Support- The system will be extended to include support for multiple languages by integrating advanced multilingual speech recognition and NLP models, allowing interviews to occur across cultural and linguistic borders.
- Cloud Deployment- Migrating the system to a cloud infrastructure such as AWS or Google cloud will improve scalability, allow for remote access to the system from anywhere and provide real-time processing capabilities for large-scale recruitment drives.
- Advanced Behavioural Analysis- Future versions of the system will utilize gesture recognition, posture-based analysis and micro-expression detection technologies in addition to facial emotion analysis to provide a more comprehensive assessment of a candidate's personality and behaviour than current systems can provide.
- Adaptive Question Generation- An AI-based, dynamic question generation module will be developed to adjust the complexity and domain of interview questions automatically based on the candidate's performance during the interview.
- HR System Integration- The recruitment platform will integrate with existing HR Management Systems (HRMS) and Applicant Tracking Systems (ATS) to provide automation of the entire recruitment process, including candidate data, end-to-end workflow and seamless integration between resourcing and HRMS/ATS.

Conclusion

The automatic and intelligent candidate evaluation system using machine intelligence shows promise based on various combinations of technology, with the goal of integrating Speech Recognition, Natural Language Processing, and Facial Emotion Detection, and Voice Biometrics into a single solution. The systems have been developed to overcome the main limitations of traditional interviews by providing unbiased, consistent, and scalable candidate assessments through the use of data-driven and objective multimodal assessments. Experimental results have demonstrated an accuracy of 92.5% for the overall evaluation of candidate performance through multimodal fusion, demonstrating the superior effectiveness of multimodal over unimodal evaluation. The inclusion of anti-cheating monitoring, automated reporting, and centralised database management enhance the reliability and integrity of the evaluation process. The architectural components of the system allow for practical application in the real world (including employment, training, or educational settings) and demonstrate scalability; hence proving the potential for future AI-based recruitment and automated evaluation systems. Overall, the technology will provide a more equitable, efficient, and intelligent means of evaluating candidates and will be a major contribution to future developments in AI-powered Human Resource Management and Automated Evaluation Technologies.

References

[1]. Joshi, K., Bachhav, M., Raut, A. and Keswad, M., 2026. AI-POWERED VOICE-BASED INTERVIEW SIMULATION SYSTEMS.
[2]. Deshmukh, M.B., Yadav, S., Kare, P. and

Khaladkar, O., 2026. AI-Driven Mock Interview: A New Era In Candidate Preparation. JOURNAL OF ADVANCE AND FUTURE RESEARCH, 4(1), pp.432-437.

[3]. Inam, S.A., Kabeer, A., Abbasi, M.A. and Khan, A.A., 2026. IVAS: A multimodal AI system for objective video interview assessment with facial emotion, gaze, and audio analysis. Edelweiss Applied Science and Technology, 10(1), pp.525-543.

[4]. Patil, P.N. and Wagh, S.J., AI-Driven Mock Interview Systems: A Comprehensive Survey of Emotion and Performance Evaluation Platforms. JOURNAL OF TECHNICAL EDUCATION, p.76.

[5]. Alkhayat, A.H., 2026. The Effect of Multimodal Conversational AI on Job Interview Anxiety and Performance among ESL Students (Doctoral dissertation, Teachers College, Columbia University).

[6]. Wen, Y., Zhang, Y., Suresh, S., Lu, Z., Liu, C. and Xia, M., 2026. InterFlow: Designing Unobtrusive AI to Empower Interviewers in Semi-Structured Interviews. arXiv preprint arXiv:2602.06396.

[7]. Handignoor, M.S., Singh, M.H., Suryawanshi, V. and Kadve, D., 2026. Artificial Intelligence–Based Mock Interviews for Performance Improvement.

[8]. Sakib, M.N., Rayasam, N.M. and Dey, S., 2026. Experience and Adaptation in AI-mediated Hiring Systems: A Combined Analysis of Online Discourse and Interface Design. arXiv preprint arXiv:2601.02775.

[9]. Habib, M., Ellis, J., Palachi, A., Korman, M.B., Kay, T., Barlow, K., DeSouza, J., Macri, R., Alabousi, A. and Anvari, M., 2026. Patient Perspectives on AI-Powered Medical Robots in Breast and Prostate Cancer Care: Qualitative Study. JMIR cancer, 12, p.e69710.

[10]. Daryanto, T.H., 2026. Towards Human-AI Teaming for Skill Development: From Dyadic Interview Practice to Triadic Programming Collaboration (Doctoral dissertation, Virginia Tech).

[11]. Muhammad, R., 2026. Qualitative Data Analysis With Chatgpt And Qualcoder: A Step-By-Step Guide To AI-Powered Coding And Thematic Analysis. Rafiq Muhammad.

[12]. Niroula, N. and Sharma, G., 2026. Opportunities, Risks and Ethics in Rising AI-Powered Hiring. The Voice of Creative Research, 8(1), pp.90-109.

[13]. Surendra, H.H., Archana, H.R., Jyothi, A.P. and Kashi, N.R., 2026. Intelligent framework for autonomous recruitment system using multi-modal attributes. Journal of Ambient Intelligence and Humanized Computing, 17(1), pp.117-124.

[14]. Sima, H., Chen, J., Wei, J., Chen, W. and Thaichon, P., 2026. Fight Fire With Fire: How Does AI-Powered Technology Empower the Elderly Anti-AI Fraud Through a Socio-Technical Systems Theory Lens?. Journal of Consumer Behaviour.

[15]. Hu, Y., Wang, Y., Watanabe, B., Gao, M., Qin, Z., Shi, K., Zhang, Z., Shambanna, S., Xue, Z., Zou, T. and Xiong, Q., 2026. " A Safe First Step": Design and Evaluation of an Emotionally Expressive AI Virtual Patient for Clinical Simulation in Speech-Language Pathology Training.