# Study on AI-Based Interview Performance, Response Analysis, and Candidate Evaluation Using Machine Learning and Natural Language Processing

Ramesh H R[1], Muktha Nayak[2], Praveen[3], B K Chandan[4], Rohith[5]
[1]*Head of the Department of Artificial Intelligence and Data Science Coorg Institute of Technology Karnataka, India*
[2,3,4,5]*Department of Artificial Intelligence and Data Science Coorg Institute of Technology Karnataka, India*

***Abstract*—Recruiters get more applications than they can realistically read, and candidates seldom hear anything useful about why their resume or interview answers fell short. Face-to-face interviews add another problem: they lean on the interviewer's personal judgement and rarely leave the candidate with specific, individual feedback. In this paper we present IntelliView, a context-aware multimodal framework for adaptive interview assessment with explainable feedback. It reads a candidate's resume with natural language processing, builds interview questions around it, and then examines the text, audio and video of each answer, looking at what was said, how it was said, and a few behavioral cues. The results are pulled together into a report that gives an overall score, separate scores for technical knowledge, communication, confidence and problem solving, along with strengths, weak spots and tailored suggestions. We built a web prototype with secure login, a dashboard that tracks a candidate's scores across attempts, and a page for past reports. IntelliView is meant for interview practice, not for making hiring decisions on its own. We describe the architecture, review related work, discuss limitations, ethics and score variability, and lay out how we plan to validate the system experimentally.**

***Index Terms*—interview assessment, mock interview, natural language processing, multimodal analysis, resume screening, explainable feedback, machine learning**

## I. INTRODUCTION

Interviews are a key stage of recruitment. Candidates are judged on technical knowledge, communication, the quality of their answers, how confident they appear, and their conduct overall. Because the judging is usually done by people, it can be subjective, and candidates rarely get detailed feedback that is specific to them.

Advances in artificial intelligence have led to automated systems that assess candidate performance in different ways. Multimodal approaches combine video, audio and text to study facial expressions, speech, engagement, speaking rate, eye contact and similar behavioral indicators.

Others have looked at gaze, posture, speech-based confidence, similarity between answers, and interactive interview agents. Resume-based question customization and domain-specific mock interviews have been explored as well, to make practice feel more relevant.

Even so, most existing systems concentrate on one part of the problem. They tend to be only loosely connected to what is known about the candidate, to the interview itself, and to the analysis that follows. Candidates also often get a single overall verdict with little explanation of what pushed the result up or down. What is missing is an approach that understands who the candidate is, adapts its questions, listens across several channels, and then explains itself.

We address this gap with IntelliView, a context-aware multimodal framework for adaptive interview

assessment and explainable feedback. It uses the resume to establish who the candidate is and to generate questions that fit. As the interview proceeds, the audio, video and text of each response are analyzed for content, speech characteristics, facial expression indicators and a small set of behavioral cues, and these are combined into a structured assessment.

The framework also explains its assessment by naming the candidate's strengths, the areas that need work, and personalized recommendations. We do not intend it to act as an autonomous hiring decision system. Its purpose is to give candidates a consistent, interactive and personalized place to practice and to see how they are doing.

This work makes four contributions:
(i) a framework that links resume context, adaptive questioning and multimodal response analysis in a single workflow;
(ii) a dimension-wise scoring scheme with an explainable feedback report;
(iii) a web-based prototype with a progress dashboard; and
(iv) a discussion of limitations, ethical considerations and a plan for experimental validation.

Section II reviews related work,
Section III describes the system architecture,
Section IV details the processing modules,
Section V discusses the prototype outputs,
Section VI covers limitations and ethical considerations,
Section VII outlines the evaluation plan, and Section VIII concludes.

## II. RELATED WORK

### A. Multimodal Interview Assessment

Several studies use more than one channel to analyze interviews. Multimodal behavioral analytics have been used to assess interview performance and generate feedback [1], and multimodal behavior analysis has been applied in job interview training systems [2]. Work on the emotional analysis of candidates in online interviews [10] and on real-time facial emotion recognition for interview feedback and career recommendations [13] pushes this further, with a focus on facial and affective cues.

### B. Mock Interview and Feedback Systems

Mock interview platforms try to give candidates a place to practice and get feedback tailored to them. Existing work pairs natural language processing with speech analysis to produce feedback [4], classifies emotion and confidence to evaluate interviews [3], and uses language processing to simulate the interview itself [12], [15]. Adaptive, realistic practice has also been explored [14]. Most of these systems stress a single aspect, such as what is said or how it sounds, instead of tying resume context to several modalities.

### C. Resume Analysis and Job Matching

Research centred on resumes uses language processing and machine learning to rank resumes and match them to jobs [5], [11], and to classify resumes from profile descriptions [6]. This helps with screening, but it is seldom linked to the candidate's later interview practice. That link is what IntelliView tries to supply.

### D. Emotion Recognition and Fairness

Speech emotion recognition [7] and facial emotion recognition with deep convolutional networks [8] supply the technical basis for behavioral indicators. At the same time, using artificial intelligence in hiring raises concerns about bias, and explainable methods have been proposed to reduce it [9]. For this reason we treat facial and speech indicators only as supporting signals and put explanations next to every score.

## III. SYSTEM ARCHITECTURE

IntelliView is built as a pipeline of modules, shown in Fig. 1. The candidate uploads a resume and then sits an interview using a microphone and webcam. The resume processing module pulls out the candidate's context, and the question generator combines it with the interview domain the candidate has chosen. Answers are passed to the text, audio and video modules, and their outputs are fused into a multimodal assessment and an explainable feedback report. Reports and scores are stored so they can be viewed again on the candidate dashboard.

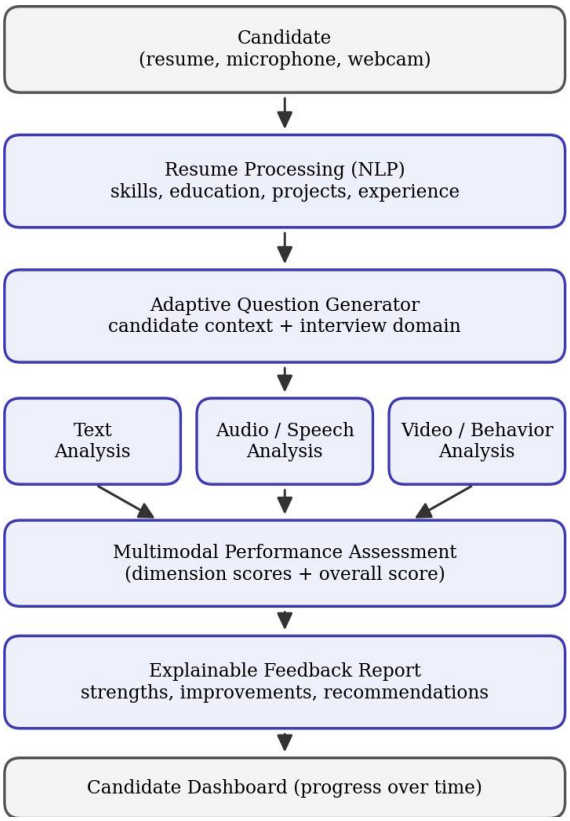



Fig. 1. Architecture of the proposed IntelliView framework

The web-based prototype is a candidate portal with five areas: Dashboard, Resume & ATS, Jobs, Interview and Reports. Resume & ATS gives a resume score, Interview runs the mock interview, and Reports keeps the generated reports. Only registered users can get in, through an email and password login.

The design is deliberately modular. Every analysis module produces its own indicators, and the fusion stage combines them so that no single indicator decides the outcome. A module can therefore be improved, swapped or switched off, for instance when a candidate prefers not to use the webcam, without redesigning the rest of the system.

## IV. EXPERIMENTAL DETAILS

IntelliView is a multimodal interview assessment system that brings together candidate context, adaptive questioning, and the analysis of audio, video and text responses. We describe it here as a set of processing modules, each covering a different aspect of interview performance.

### A. Candidate Profile and Resume Processing

The system takes candidate details and resume content as its first input. Natural language processing is applied to the resume to pick out technical skills, educational background, projects and experience. This information forms the candidate context and feeds into question generation. Access is limited to registered users through the email and password login screen in Fig. 2.

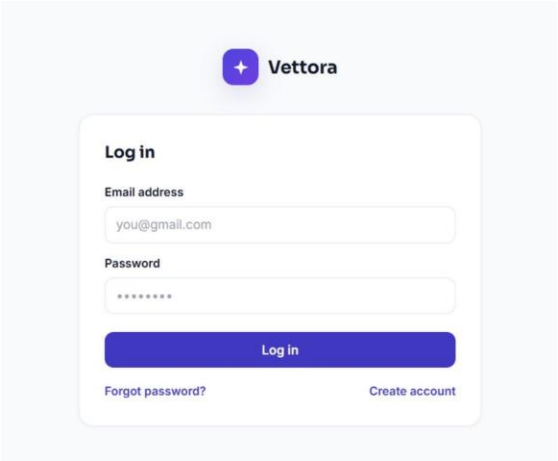



Fig. 2. User login interface of the prototype

### B. Adaptive Interview Question Generation

An AI-driven interviewer writes questions from the candidate context and the chosen interview domain. Both technical and behavioral questions are supported, and the candidate's earlier answers decide which follow-up questions come next. This makes the interview more personal than one that runs through a fixed list of questions.

### C. Audio and Speech Analysis

While the candidate speaks, a microphone captures the audio and speech recognition turns it into text. The audio can also be processed for indicators such as speaking rate, pauses and other acoustic features. We use these as supporting information when judging communication and fluency.

### D. Video and Behavioral Analysis

The webcam stream is processed with computer vision techniques, and facial and selected behavioral indicators are analyzed during the interview. We treat this visual information as supporting evidence of how the candidate behaved, not as a definitive measure of what they feel inside.

E. Textual Response Analysis

Transcribed answers are analyzed with natural language processing for relevance, clarity and how well they address the question asked. This text analysis is then combined with the audio and video information to give a fuller picture of how the candidate performed.

F. Multimodal Performance Assessment

The text, audio and video outputs are combined into a structured interview assessment. Several performance dimensions are considered, so no single characteristic dominates. The final assessment is meant to summarize what can be observed in the interview and to point out where the candidate needs to improve. Table I lists the four performance dimensions reported by the prototype and the kinds of evidence behind each.

TABLE I. PERFORMANCE DIMENSIONS AND SUPPORTING EVIDENCE

| Dimension | Main evidence | Supporting evidence |
|---|---|---|
| Technical | Correctness and depth of answers | Relevance to resume context |
| Communication | Clarity and relevance of answers | Speaking rate and pauses |
| Confidence | Speech pattern indicators | Selected visual behavioral cues |
| Problem solving | Structure of reasoning in answers | Follow-up question responses |

G. Feedback Generation

From the assessment results, IntelliView produces an explainable performance report with strengths, areas for improvement and personalized recommendations. Feedback is tied to the performance indicators that were analyzed, so candidates can see why they were assessed the way they were and use that when preparing further. Fig. 3 shows an example report, with an overall score, four-dimension scores (technical, communication, confidence and problem solving) and an AI recommendation.




Fig. 3. Interview report generated by the prototype

H. Score Reporting

The prototype reports each dimension and the overall result on a 0–100 scale. The overall score is derived from the dimension scores, but it need not be their plain average, since the dimensions can carry different weights. The weighting directly shapes what candidates are told, so the aggregation scheme should be documented and checked against human ratings, as described in Section VII.

## V. RESULTS AND DISCUSSION

IntelliView offers a single environment for running and analyzing simulated interviews using candidate context and multimodal information. It is designed to put textual response analysis alongside audio and visual indicators, so the assessment covers more than the words alone.

The resume processing module identifies details such as skills, education, projects and experience, which give context for generating relevant questions. Adaptive questioning lets later questions be shaped by what the candidate said earlier, which makes the interview feel more personal than a fixed sequence.

The multimodal analysis component handles the text, audio and video collected during the interview. Text analysis tells us how relevant and clear the responses are, audio processing looks at fluency and speaking patterns, and video processing adds supporting information on selected facial and behavioral indicators. Using the three together gives a fuller view of interview performance than any one source.

The performance report gives structured feedback as strengths, areas needing improvement and personalized recommendations. The idea is to help candidates understand how they actually performed and what to work on next. The candidate dashboard (Fig. 4) shows the resume ATS score, the latest interview score and the number of interviews completed, and plots performance over time so candidates can follow their progress from one attempt to the next.

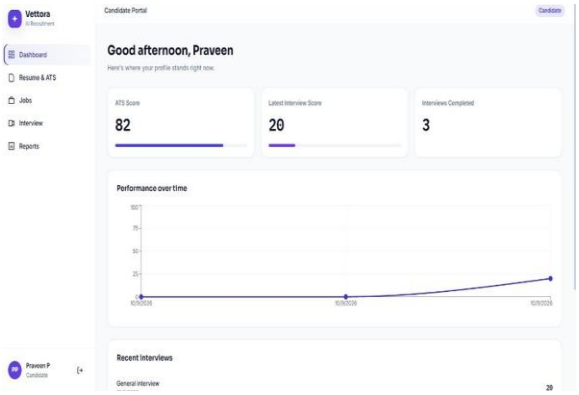



Fig. 4. Candidate dashboard showing ATS score, latest interview score and performance over time

The report and the dashboard do different jobs. The report explains one attempt, while the dashboard shows the trend across attempts, which matters most when someone practices repeatedly. With both, the explanation of individual scores is connected to longer-term improvement.

That said, how well the framework works depends on the quality and diversity of the data, the reliability of each analysis module and the robustness of multimodal integration. Facial and speech indicators should therefore be read as supporting behavioral signals, not as definitive measures of a candidate's emotions or personality. A larger and more diverse dataset is needed to validate the framework's performance quantitatively.

## VI. LIMITATIONS AND ETHICAL CONSIDERATIONS

Model variability. Scores that rely on language models or on speech and vision components can change between runs, question sets and recording conditions. They should therefore be presented as practice feedback, and repeated runs on the same answers should be checked for consistency.

Bias and fairness. Speech and facial indicators can be affected by accent, language background, speaking style, disability, lighting and camera quality, none of which say anything about a candidate's ability. These factors can put some groups at a disadvantage, so every module should be tested for fairness across groups before its scores are relied on [9].

Privacy and consent. Interviews involve audio, video and personal resume data. Candidates should be told what is captured, should consent before the microphone or webcam is switched on, and should be able to turn off the video module. Recordings and transcripts should be kept only as long as needed and protected by access control.

IntelliView is meant for interview practice and self-assessment. Its outputs, including the recommendation label in the report, should never be the sole basis for a hiring decision, and a human reviewer should stay responsible for any selection outcome.

Scope of evidence. The prototype outputs in this paper illustrate the interface and report structure. They are not experimental results and do not establish the accuracy of the framework.

## VII. PROPOSED EVALUATION PLAN

To validate the framework quantitatively, we plan to evaluate the aspects summarized in Table II. Mock interviews will be collected from volunteer candidates who have given consent, and experienced human reviewers will rate each one independently on the same four dimensions. We will then measure how closely the automated scores agree with the human ones.

TABLE II. PLANNED EVALUATION ASPECTS

| Aspect | Planned measure | Purpose |
|---|---|---|
| Agreement with reviewers | Correlation and weighted kappa | Score validity |
| Consistency | Score variation over repeated runs | Reliability |
| Fairness | Score differences across groups | Bias check |
| Usefulness | User survey and score trend over attempts | Practice benefit |
| Usability | Standard usability questionnaire | Interface quality |
| Responsiveness | Time to produce a report | Practical use |

Each analysis module will also be tested separately, for example the accuracy of speech transcription and of the textual relevance judgement, so that an error can be traced to the module behind it. Ablation experiments that remove one modality at a time will show how much text, audio and video each add to the final assessment. We will report results only once these experiments are complete.

## VIII. CONCLUSION

This paper presented IntelliView, a context-aware multimodal framework for adaptive interview assessment and explainable performance feedback. It brings together resume information, adaptive question generation, and the analysis of text, audio and video responses to give a structured view of how an interview went. By drawing on several sources of information at once, the framework aims to overcome the limits of approaches that look at only one assessment component.

It also gives personalized feedback, naming strengths, areas for improvement and recommendations based on the observed performance indicators. We hope this helps candidates prepare better while keeping the assessment process transparent. Future work will focus on thorough experimental validation with diverse datasets, better multimodal analysis, evaluation of fairness and robustness, and optimizing the framework for real-time interview assessment.

## ACKNOWLEDGMENT

The authors sincerely thank the faculty and management of Coorg Institute of Technology, Karnataka, for their guidance, support and encouragement throughout this research. We are also grateful to our project guides and mentors for their continuous suggestions and technical help in designing and developing IntelliView. Finally, we thank our colleagues and everyone who contributed, directly or indirectly, to completing this work.

## REFERENCES

[1] Agrawal, R. A. George, S. S. Ravi, S. Kamath, and K. M. Anand, “Leveraging multimodal behavioral analytics for automated job interview performance assessment and feedback,” 2020.

[2] N. Takeuchi and T. Koda, “Initial assessment of job interview training system using multimodal behavior analysis,” in Proc. 9th Int. Conf. Human-Agent Interaction, 2021.

[3] N. S. Rai, A. K. R., A. P., and H. N. R., “AI based interview evaluator: An emotion and confidence classifier,” Int. J. Adv. Res. Comput. Commun. Eng., vol. 13, no. 4, pp. 747–757, 2024.

[4] T. K., A. T. V., K. K. R., R. B. E., and S. M. K., “AI-powered mock interview platform with NLP and speech analysis for personalized feedback,” Int. Res. J. Adv. Eng. Hub, vol. 3, no. 8, pp. 3335–3339, 2025.

[5] “Resume matching framework via ranking and sorting using NLP and deep learning,” 2023.

[6] “Real-time resume classification system using LinkedIn profile descriptions,” 2020.

[7] “A survey on speech emotion recognition,” 2019.

[8] “Real-time facial emotion recognition using deep convolutional neural networks,” 2023.

[9] “Mitigating bias in AI model using explainable AI in terms of hiring process in the industry,” 2025.

[10] “Emotional analysis of candidates during online interviews,” 2023.

[11] “Enhanced algorithmic job matching based on a comprehensive candidate profile using NLP and machine learning,” 2023.

[12] “AI based mock interview system using natural language processing,” in IEEE, 2025.

[13] “AI interviews with facial emotion recognition for real-time feedback and career recommendations,” in IEEE, 2025.

[14] “MockMate: Leveraging AI for adaptive and realistic interview practice,” in IEEE, 2025.

[15] “AI-powered virtual job interview simulator using natural language processing,” in IEEE, 2025.