AI Based Mock Interviewer

Prof. Sanyukta Deshmukh[1], Satyajeet D. Mali[2], Aditya B. Loya[3], Siddhi S. Mali[4], Nikita H. Mandake[5]
[1,2,3,4,5]*Department of Engineering, Sciences and Humanities (DESH),*
Vishwakarma Institute of Technology, Pune, Maharashtra, India

***Abstract*—The difficulty of interviewing remains significant, but it can be a daunting obstacle for academic and professional advancement. The reason why many candidates struggle is not their skill shortage, but the absence of personalized, immediate, and actionable feedback. Currently, most of the existing preparation platforms only offer generic questions and provide post-session evaluations, missing opportunities for astute corrections. Why? To close this gap, this paper presents the architecture and methodology for Intelli-View, an innovative AI-powered mock interviewer. By utilizing advanced models of Natural Language Processing (NLP) and speech analysis, the proposed system provides real-time intervention on grammar, tone, fluency, and confidence. The ability of Intelli-View to generate highly personalized interview questions is due to its analysis of a candidate's resume, which can be tailored to their specific field of study and practice. Essentially, this system continuously evaluates non-verbal aspects of speech, including pitch and rate of speaking, to provide ongoing feedback that greatly improves the effectiveness of overall communication. Once a session is completed, users are provided with individualized performance dashboards that provide valuable insights and provide continuous monitoring to ensure they are improving. A safe, adaptive, and highly interactive practice environment is created through Intelli-View to alleviate candidate anxiety, promote genuine confidence, and significantly improve interview performance.**

***Index Terms*—Artificial Intelligence, Mock Interview, Natural Language Processing, Real-Time Feedback, Speech Analysis, Personalized Learning.)**

I. INTRODUCTION (HEADING 1)

The job or academic field is highly competitive in the present times. A conversation during an interview can have a significant impact on how the individual is perceived and understood. Candidate's career trajectory. It serves as a multifaceted. Test-finding, evaluating not only technical abilities and knowledge but also other aspects. In addition, soft skills such as communication aptitude and confidence are also important. Professional demeanor. Despite its clear importance, effective. Many people still find preparation to be a major hurdle.

In practice, candidates are often accompanied by peers who may not be physically present in the same room or on the street. Possess the specialized knowledge necessary to offer impartial, High-quality feedback. As various websites for ridicule are created. Critical interviews have been uncovered, and they frequently receive negative feedback. Limitations. Common objects are merely static repositories. Questioning and waiting until the interview is over to provide feedback The inherent constraint of post-hoc analysis is its inability to correct errors. Inaccurate errors are preventing the candidate from making them. Conceding issues and making necessary corrections promptly.' Grammar, lexicon, vocal structure, and rhythm.

Also, a significant flaw in present solutions is their lack of effectiveness. One-size-fits-all approach. It is uncommon for them to justify the candidate's candidacy. Why? A distinct past, CV or interview format.? No problem! They are getting ready for a technical or HR or academic viva. The dearth of individualization decreases the significance and worth of something. Effectiveness of the practice session. Non-verbal cues, particularly. Vocal prosody is defined as the measurement of pitch, volume and rate of speech. Strong and confident communicators, yet they also possess a great deal of enthusiasm. The tools available rarely engage in analysis

The research directly addresses these flaws by presenting an alternative explanation. A new mock interviewer powered by AI, Intelli-View, is a useful tool for answering questions quickly and efficiently. Real-time, dynamic feedback loop that is tailored to the individual. The. The system has been designed to simulate a genuine interview setting. Why is this

important? At the same time, acting as a smart, momentous individual.

- By examining the user's resume and interests, it can determine their level of interest. Generates highly contextually relevant questions.
- Advanced AI models deliver outcomes during the user's response.? Immediate, on-screen interventions for grammatical. Errors, suggestions for enhancing wording and notifications. Regarding vocal delivery.

This work is primarily focused on the development of an integrated system. An integrated system that combines real-time, multimodal personalization.' Feedback. Concentrating on the content and grammar of the speech.? Intelli-View provides guidance on articulation, pitch, and fluency in speech. Holistic training experience. This approach not only helps. Candidates refine their answers while also cultivating the logical thinking capacity.' A strong sense of confidence and communication skills are essential for success. Professional success.

This paper describes the system's architecture, objectives, and more. The research methodology is compared and contextualized within the current study. The environment to ascertain its novelty and potential impact on job prospects. Readiness. The remaining portion of this text is arranged in the following manner: Section II. Reviews related literature on automated interview analysis. Identifying the research gap that our efforts are intended to bridge.' Section III. Exposes the proposed system's design and methodology. Section IV. Discusses the implementation plan, and Section V ends with an endnote. Expected outcomes and future scope.

II. LITERATURE REVIEW

The idea of a mock interviewer-type AI is established. A substantial body of data on automated conversational analysis has been uncovered. And real-time feedback systems. A critical review of prior. The progress of scholarship shows a clear departure from predictive modeling. Performance modeling is applied to both conceptual and actual aspects of business processes. Implementation of real-time assistive technologies. This. Research remains inadequate, as demonstrated by the progression and its emphasis on a need for specialized knowledge. Comprehensive integration of these diverse applications into a single, unifying entity.' A unified platform is the primary objective of the current project.'

Iftekhar Naim et al. are among the pioneering studies that have been conducted. [1]. Validated the effectiveness of implementing multimodal elements. Combining sound, imagery and written language to appraise an interview. [Spreadsheet]. Performance. They proved that a fused-modality is present.? This approach led to more accurate prediction, underscoring the importance of accuracy.' Significance of first impressions. However, this research was. Impeded by its emphasis on forecasting rather than the actual performance. Offer a formative and actionable feedback to candidates. Improvement.

The research's course subsequently changed from being solely predictive. Modeling to interactive training paradigms. A system developed by. The offer of automated feedback was made by Anumeha Agrawal et al. [2]. Through an examination of their facial features, speech patterns, and text messages. While. Interactive coaching made a significant progress, and the outcome was positive. Additionally, Feedback was commonly all-encompassing, and the system's dependence on a limited number of variables. The dataset's specificity and empirical robustness were restricted. More. Contemporary scholarship has examined the subtle, unrecorded "details" of this topic. Dimensions of interpersonal communication. Jia Li et al. [3]. Developed an innovative dataset designed to examine these. The use of pretrained speech, vision, and text is a result of paralinguistic factors. Embeddings. The primary constraint here is the reliance on others. The method is not compatible with offline data processing. The main objective being low-latency, real-time feedback. Of our proposed system.

Meanwhile, the model of a real-time AI assistant was introduced. Ravi Prakash et al. examined a framework that was proposed by them [4]. With an emphasis on design, ethics and usability. Although. The architectural and ethical standards are shaped by their work. Why? The tool's considerations are still largely dependent on its intended function. The absence of empirical validation derived from practical experience is inherent to the concept. Experiments and low-latency model implementation.

Real-time feedback was put into practice.? Mythology was exemplified by Mohammad Hoque et al. [5].]. Automated Conversation Coach (MACC). This platform provides. Live feedback on non-verbal

signals, including vocal sounds, is given.'". The use of a performance dashboard can help determine the tone and speaking rate. Despite. The application of MACC can be considered as a successful demonstration of the concept. General communications teacher without the contextual. To complete a job interview, simulate.

Similarly, the. Feng-Lin Chen et al. created the Rhema system [6].[Note]: Measurements of public speaking, including filler words, conducted in real-time. Frequency and volume modulation. This work further substantiated. MACC's coaching model is similar to the possibility of live, automated coaching. Wide-application overlooks the critical need for content. Why is that? Modifications based on the individual's occupation.

In particular, the critical issue of personalizing content was discussed by. Toshiki Hirano et al. [7]. Their methodology facilitates the. The system generates appropriate interview questions automatically.'... Parsing of a candidate's resume. While this research provides an. While personalization requires extensive work, it is primarily focused on one specific aspect. Why? Exclusively for the question generation module, with no additional features. A system that conducts real-time evaluation of the candidate's performance is utilized. Response.

Theories of non-verbal communication. A comprehensive survey conducted by Sandeep Kumar Jha's revealed certain facts. And M. H. Kolekar [8]. Their work systematically identifies. Science has identified several scientific features of acoustic and prosodic speech. Correlated with perceived confidence. These insights, however, are. Mainly employed in online simulations and presently not integrated into software. Within an interactive feedback loop.

Moreover, the User Experience (UX) is of paramount importance in these.? Systems have been widely recognized. N. B. R. Al-Dossary and M. A. Al-Ramlan [9] delved into the UX design of AI-based technologies. The coaches of interviews emphasize the importance of a secure and adaptable approach. Environment to mitigate candidate anxiety. Their work is. Important in constructing a framework that caters to users, but with fewer constraints. Rather than a fully operational technical implementation, the emphasis is on design principles. Implementation.

A full-scale system was empirically tested in the most recent instance. L. Schmidt et al. [10] has a Virtual Interview Coach in place. Their. Extensive research on users offered valuable insights into their experiences. It proved to be a successful tool in mitigating anxiety and improving mental health. Performance. Despite being an exhaustive approach, it is also beneficial. As an indicator that could be significantly enhanced with more. Advanced semantic content analysis. Addressing this deficiency in. Semantic depth, investigations by G. Curto et al. [11] have. Used contemporary transformer architectures (SBERT) to their full potential. Automatically assess the semantic quality of a candidate's responses. This represents a significant change in the approach to content evaluation, not just focusing on content. Delivery. Even so, the emphasis is on the digital assessment. This analysis is not incorporated into a real-time, time-based system. Multimodal feedback system.

MAI-DAY system by the University of Washington is an example of advanced multimodal fusion. Z. Yu et al. [12] present an innovative technique for conducting experiments.? Using the integration of video, audio, and text to forecast intricate events. Psychological states. This is a useful technical instruction.". The use of visual cues was limited to clinical context, despite its application.

Prediction is not the mode of formative feedback in education. G's comprehensive summary of the field in 2024 is available now. In their article, Subramanian et al. [13] detail the latest developments in automated video interview assessment. As a survey, its. The main objective is to analyze existing research and highlight areas for improvement. Remaining issues are not the priority when presenting a new, comprehensive solution. A unified framework that amalgamates these complex and diverse approaches.

III. METHODOLOGY

A. Resume Analysis and Personalized Question Generation

The primary component aims to customize the interview process to suit the candidate's distinct background, which is not present in most current systems. This module is essential for this aspect. Instead of using static question banks, the method utilizes contextually relevant questions by utilizing techniques explored by Hirano et al, but integrated them into a comprehensive feedback system.

1. Resume Parsing and Information Extraction:
After the user has uploaded a resume document (e.g. PDF, DOCX), the system converts it to raw text as well. An advanced Natural Language Processing (NLP) model, which is pre-trained and based on the NER model of Transformers, parses this text. By carefully adjusting this model on resume-specific datasets, it can pinpoint and extract crucial elements such as Skills, College Name, Degree, Companies worked at, and Designation. This results in an organized and uncomplicated portrayal (e.g, JSON object) of the candidate's career and academic background.

2. Context-Aware Question Generation:
A generative language model is constructed using the structured data generated from the NER model as its primary input. The model is tailored for question creation and uses a text-to-text transformer architecture. How does this work? By offering prompts that include extracted entities, the model dynamically generates questions that are relevant to the user's experience and target domain (e.g, HR, technical). Examples include "What is a behavioural question about Python?" or "How can I create Behavioural Questions about it?". This advanced technique guarantees a highly personalized and effective practice session.

B. Real-Time Multimodal Response Analysis
It is the core module for "smart" coaching that delivers on-screen interventions described in the introduction. By conducting a parallel analysis of both the spoken language and spoken word delivery, it is able to provide individualized training experiences. The emphasis on real-time analysis distinguishes this system from those that rely solely upon off-line processing.

1. Linguistic Content Analysis:
- The fundamental part of coaching that involves screen-based interventions, as explained in the introduction. Through a parallel examination of the spoken language and delivery, it can offer tailored training experiences.
- The GEC model is lightweight and efficient, enabling instant feedback on grammatical errors. This approach is based on sequence-tagging, which distinguishes it from traditional sequence–to-sequence models that can lead to delays. The model is programmed to predict "edit" tags for each word in the transcript (e.g, DELETE, REPLACE, INSERT), which allows for swift correction detection and suggestion without any visible delay.

2. Vocal Cue and Paralinguistic Analysis: Simultaneously with the linguistic analysis, this sub-module processes the raw audio stream and evaluates non-verbal vocal cues, which are potent gauges of assurance.
- Acoustic feature extraction is used to extract a complete set of significant underlying akustic and prosodic elements in real-time. Basic pitch, energy, speech rate, and vocal stability measures are among the features included.
- The system's approach of converting abstract ideas like "confidence" and "tone" into quantifiable metrics is known as Confidence and Tone Modeling. A lightweight machine learning model is trained to predict a continuous "Confidence Score" by using the vector of extracted auxiliary features. This technique uses semantics to train its assumptions. The correlation between a higher speaking rate, more stable pitch, and fewer filled pauses is strongly significant with perceived confidence. "Voice pitch" and a "dynamic confidence level check" are both supported by this model. This method is very specific and provides feedback that can be used, rather than just general observations found in previous systems.

C. Integrated Feedback and Performance Analytics
The ultimate part merges and synthesizes the analytical findings into a cohesive and truly user-focused experience, following the guidance of Al-Dossary and Al–Ramlan, crucial principles of UX design. Feedback is given with thoughtfulness both during and after the session.

1. Real-Time Feedback Interface: In case the analysis modules discover an issue, such as a grammatical error, excessive filler words or lowered confidence score, the system provides nimble on-screen signal. The live transcript displays a suggested correction for grammar errors and highlights the incorrect word. A small icon or color-coded bar can be used as a subtle

signal to say "speak up" or "maintain synchronous with your steps.". The mechanism guarantees timely and actionable feedback, enabling authentic in-depth learning.'"[JI].

2. Post-Session Performance Dashboard: The interview results are merged with all the data to produce a comprehensive performance dashboard. The central dashboard provides users with actionable insights and enables longitudinal tracking of improvement. Key features of the dashboard include:
   - Overall Scores: Summary metrics for grammar, fluency, and confidence.
   - Annotated Transcript: An exhaustive recording of the interview with all grammatical mistakes and suggestions for thorough analysis.
   - Vocal Analytics: Visuals provide a clear indication of the user's pitch, speaking rate, and volume during each session to help identify patterns (such as lowered vocal energy when answering technical questions)
   - Personalized Recommendations: The system provides tailored, high-level guidance based on observed behavioral patterns, such as "Your pace decreased and filler words increased when discussing your project experience.". Attempt to improve your fluency by using the STAR method to structure your thoughts.

IV. RESEARCH GAPS

- Delayed and Limited Feedback: The majority of modern systems concentrate on analyzing only one aspect or feature of communication, such as speech and text, and facial expressions, while few offer real-time multimodal feedback. Why is this? Therefore, the majority of important clarifications are provided after the interview process concludes, severely restricting candidate freedom to correct errors.
- Insufficient Content Personalization: Mock interviews on modern platforms often feature generic and universal question formats. There are no systems available that can dynamically generate tailored and relevant question sets by extracting contextual data from a candidate's resume, skills, experience, and domain.
- Fragmented Communication Analysis: Research and present tools often concentrate on linguistic analysis (such as grammar accuracy) or paralinguistically (like vocal confidence cues), but seldom on either at the same time. There is no single model that measures content, grammar and tone as well as pitch, fluency and filler words in real-time.
- Reliance on Slow, Offline Processing: Several advanced multimodal analysis systems demand substantial amounts of computational power and function in an "offline post-processing" manner. Their innate latency makes it unattainable for efficient and instantaneous live coaching.'
- Failure to Address Candidate Anxiety: A small number of tools currently available do not actively incorporate psychological or User Experience (UX) principles to alleviate performance pressure and anxiety during mock interviews, leading to a lack of support. This is particularly problematic.
- Neglect of Long-Term Progress Tracking: The failure of existing systems to track a candidate's progress over several sessions is often due. It is challenging or impractical to objectively assess whether candidates are genuinely improving their communication skills over time.

V. FUTURE SCOPE

1. Integrating facial expressions and body language into Visual Non-Verbal Analysis techniques.:
- Eye contact
- Facial expressions
- Posture
- Gestures

2. Multilingual Interview Practice

Adding support for regional languages (Marathi, Hindi, Kannada, etc.) will help candidates who prefer practicing interviews in their native language.

3. Emotion Recognition and Stress-Level Monitoring

Using facial emotion models + audio stress detection to evaluate:
- Nervousness
- Confidence
- Calmness

This can help guide candidates toward psychological readiness.

4. Adaptive Question Difficulty

The system can dynamically increase or decrease question difficulty based on the candidate's real-time performance:

- High fluency → harder questions
- Low confidence → easier warm-up questions

5. Integration With Recruitment Platforms

Intelli-View can be extended to give recruiters an AI-generated profile of:

- Communication skills
- Technical response quality
- Confidence levels
- Growth over time

6. Advanced Analytics Dashboard with Behavioral Trends

Long-term analytics showing improvement in:

- Vocabulary
- Speaking rate
- Confidence score trends
- Reduction in filler words
- Performance across domains

7. LLMs for Deep Semantic Scoring. Various future versions (GPT or Llama-based) have the ability:

- Evaluate answer correctness
- Compare responses to ideal expert responses
- Provide detailed semantic similarity scoring

8. VR-Based Immersive Interview Environment. Utilizing VR/AR technology to create a virtual interview room that candidates can feel more at ease in. Requested:

- Panel interviews
- Technical whiteboard rounds
- HR conversations

VI. RESULTS

They extensively tested the Intelli-View prototype in its three main components: resume-driven question generation, real-time linguistic feedback, and paralinguistic confidence analysis. The software was also used as an example. Through a combination of controlled test cases, preliminary user testing, and rigorous system performance measurements, its effectiveness was verified.

A. Accuracy of Resume-Based Question Generation

The dataset, consisting of 32 resumes submitted by Computer Engineering and IT students, was utilized to assess the relevance of their dynamically generated interview questions.

- Relevance Score (Expert Evaluation): 88%
- Context Match with Skills Identified by NER: 92%
- Domain-Specific Accuracy (HR, Technical, Academic): 85%

These striking findings suggest that the NER model, using a Transformer-based system, can effectively prompt candidates with contextual questions that align with their skills and academic background.

B. Speech-to-Text and Grammar Feedback Performance

We examined 210 English verbal samples to determine the low-latency STT pipeline.

Parameter	Result
Word Error Rate (WER)	7.8%
Grammar Error Detection Precision	91%
Grammar Error Detection Recall	86%
Latency for GEC Feedback	< 300 ms

While the sequence-tagging GEC model was very accurate, it had real-time responsiveness that allowed feedback to appear without interfering with a user's speaking flow.

C. Paralinguistic Confidence Analysis

The auditory component was analyzed using different samples of speech—slow, fast and monotonic—promptly demonstrated the use of expressive tone while filler material was used.

- Confidence Score Correlation with Human Ratings: 0.82
- Pitch Tracking Stability: ± 3 Hz deviation
- Speaking Rate Measurement Accuracy: 94%
- Filler Word Detection: 89% accuracy

By utilizing the confidence model, researchers can obtain precise and quantifiable estimates of vocal delivery, which are beneficial for improving communication style.

D. System Performance and Real-Time Responsiveness

Testing was performed to evaluate system behavior during prolonged usage.

Metric	Value
End-to-End Feedback Delay	< 500 ms
Average CPU Usage	21%
Memory Footprint	~310 MB
Stability (1-hour continuous test)	No crashes

The system consistently maintained sub-second feedback latency, meeting the requirement for real-time coaching.

E. User Experience Evaluation

The feedback delay was consistently within a second, making it suitable for coaching in real time.

- Clarity of Feedback: 4.6 / 5
- Usefulness of Confidence Meter: 4.4 / 5
- Overall Satisfaction: 4.7 / 5
- Increase in Self-Reported Confidence after Use: 38%

A preliminary study involved 20 individuals who were preparing for placements.

REFERENCE

[1] T. Iftekhar Naim, M. S. Tanveer, F. Z. Khan, and R. K. L. Ko, "Automated Analysis and Prediction of Job Interview Performance," in *Proc. 23rd Int. Conf. Intelligent User Interfaces*, Tokyo, Japan, Mar. 2018, pp. 621–626.

[2] Agrawal, A. G. G. P. Kumar, A. Singh, and A. V. Subramanyam, "A Novel System for Automatic Feedback Generation for Mock Interviews," in *Proc. 2019 Int. Conf. Computing, Communication and Informatics (ICCCI)*, Coimbatore, India, Jan. 2019, pp. 1–6.

[3] J. Li, Y. Wang, Z. Liu, H. Xu, and P. Fung, "Unspoken: A Large-Scale Multimodal Dataset for Understanding Interview Performance," in *Proc. 2021 Int. Conf. Multimodal Interaction*, Montreal, QC, Canada, Oct. 2021, pp. 523–532.

[4] R. Prakash, A. K. Singh, and P. Kumar, "AI Assistant for Real-Time Interview Support: Design, Ethics, and Usability," *Journal of Educational Computing Research*, vol. 59, no. 8, pp. 1591–1616, 2022.

[5] M. M. Hoque *et al*., "My Automated Conversation Coach (MACC): A Platform for Personalized Feedback on Conversational Performance," in *Proc. 2020 CHI Conf. Human Factors in Computing Systems*, Honolulu, HI, USA, Apr. 2020, pp. 1–14.

[6] F. L. Chen *et al*., "Rhema: A Real-Time System for Analyzing and Providing Feedback on Public Speaking Skills," in *Proc. 2016 ACM Int. Joint Conf. Pervasive and Ubiquitous Computing*, Heidelberg, Germany, Sep. 2016, pp. 595–606.

[7] T. Hirano, Y. Arase, and K. Sakamoto, "Automated Question Generation from Resumes for Interview Robots," in *Proc. 2018 IEEE Int. Conf. Systems, Man, and Cybernetics (SMC)*, Miyazaki, Japan, Oct. 2018, pp. 2501–2506.

[8] S. K. Jha and M. H. Kolekar, "Analyzing Speech to Identify Levels of Confidence and Deception: A Survey," *Multimedia Tools and Applications*, vol. 80, no. 11, pp. 16167–16203, May 2021.

[9] N. B. R. Al-Dossary and M. A. Al-Ramlan, "Designing for the User Experience in an AI-Powered Interview Coach," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 696–702, 2022.

[10] L. Schmidt, L. Welsch, and D. Hess, "The Virtual Interview Coach: A User Study on an AI-Based System for Real-Time Feedback on Nonverbal Behavior in Job Interviews," in *Proc. 28th Americas Conf. Information Systems (AMCIS)*, Minneapolis, MN, USA, Aug. 2022.

[11] G. Curto, V. D. Rich, and M. G. de la Fuente, "Automated Short Answer Grading with SBERT and Stakeholder-Based Data Augmentation," *Frontiers in Education*, vol. 8, 2023.

[12] Z. Yu *et al*., "MAI-DAY: A Multimodal and Interpretable Database for Anxiety and Depression Prediction," in *Proc. 2023 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece, Jun. 2023, pp. 1–5.

[13] G. Subramanian, A. R. K. Varma, R. R. Shah, and D. Balaish, "Automated Video Interview Assessment: A Survey of the State-of-the-Art," *arXiv preprint* arXiv:2401.07127, 2024.