PREVUE.AI: A Web-Based Intelligent Mock Interview System Using Speech and Non-Verbal Analysis

Amal Joy
Dept. of Computer Science and Engineering
St. Joseph's College Of Engineering And Technology, Palai
Kerala,India
amaljoy2026@cs.sjcetpalai.ac.in

Anush S Kumar
Dept. of Computer Science and Engineering
St. Joseph's College Of Engineering And Technology, Palai
Kerala,India
anushskumar2026@cs.sjcetpalai.ac.in

Bijal T Benny
Dept. of Computer Science and Engineering
St. Joseph's College Of Engineering And Technology, Palai
Kerala,India
bijaltbenny2026@cs.sjcetpalai.ac.in

Jismi Saju
Dept. of Computer Science and Engineering
St. Joseph's College Of Engineering And Technology, Palai
Kerala,India
jismisaji2026@cs.sjcetpalai.ac.in

Thushara Sukumar
Dept. of Computer Science and Engineering
St. Joseph's College Of Engineering And Technology, Palai
Kerala,India
thusharasukumar@sjcetpalai.ac.in

***Abstract*—Interview preparation is a critical yet challenging phase for job seekers, as conventional preparation methods often lack realism, consistency, and objective evaluation. Traditional mock interviews depend heavily on human availability and subjective feedback, making them difficult to scale and standardize. To address these challenges, this paper presents PREVUE.AI, an intelligent mock interview platform designed to simulate technical and HR interview scenarios using artificial intelligence. The proposed system generates adaptive interview questions based on user responses, enabling a dynamic and context-aware interview flow. Spoken responses are captured and converted into text using automated speech recognition, allowing structured evaluation of answer relevance, clarity, and completeness. In addition to verbal analysis, the system incorporates basic non-verbal observation using MediaPipe-based facial landmark tracking to monitor visual indicators such as face presence and eye contact during interviews. This approach enhances interview realism without relying on complex emotional or sentiment detection. Interview results, scores, and session metadata are stored and visualized through an interactive dashboard, enabling users to track performance trends and improvement over multiple sessions. The proposed platform aims to provide an accessible, scalable, and objective interview preparation solution that closely reflects real-world interview conditions. By integrating adaptive questioning, speech-based evaluation, MediaPipe-driven visual observation, and performance analytics within a unified framework, PREVUE.AI supports continuous learning and effective interview readiness.**

***Index Terms*—AI Interview, Mock interview system, Speech-to-text, MediaPipe, Interview analytics, Performance evaluation,Speech Recognition**

I. INTRODUCTION

Interview performance plays a crucial role in the recruitment process and often determines a candidate's success in securing employment. Despite possessing strong technical knowledge or academic qualifications, many candidates struggle during interviews due to insufficient practice, lack of structured feedback, and limited exposure to realistic interview environments. Traditional interview preparation methods, such as peer-based mock interviews or static question banks, are often subjective, inconsistent, and difficult to scale, resulting in uneven preparation outcomes.

Recent advancements in artificial intelligence have enabled the development of automated interview preparation systems that simulate real-world interview scenarios. AI-driven platforms utilize Natural Language Processing and large language models to dynamically generate interview questions and evaluate candidate responses in a consistent and objective manner. Such systems reduce dependency on human interviewers while improving accessibility and scalability, making interview preparation more widely available and cost-effective [2].

In addition to verbal responses, non-verbal presentation significantly influences interview performance. Visual cues such as face visibility and head orientation contribute to perceived confidence and engagement during interviews. Modern web-based computer vision frameworks allow basic non-verbal observation without requiring specialized hardware or

heavy computational resources. MediaPipe enables efficient facial landmark tracking directly within browser-based environments, making it suitable for scalable web applications.

Motivated by these developments, this paper presents PREVUE.AI, an AI-driven mock interview platform built using the MERN stack. The proposed system integrates adaptive interview question generation, automated speech-based response evaluation, MediaPipe-based non-verbal observation, and performance analytics within a unified web application. The primary objective of PREVUE.AI is to provide a realistic, scalable, and structured interview preparation solution that supports continuous performance improvement and better interview readiness.

II. RELATED WORK

Research on AI-based interview preparation systems has gained increasing attention in recent years due to the growing demand for scalable and objective interview training solutions. Early mock interview platforms primarily relied on Natural Language Processing techniques to analyze textual interview responses and generate automated feedback. These approaches evaluated response relevance, explanation depth, and linguistic structure, demonstrating that NLP-driven analysis could assist candidates in improving interview performance without continuous human intervention [1]. However, such methods were largely limited to text-based evaluation and lacked real-time interaction and comprehensive performance tracking.

With advancements in artificial intelligence, adaptive mock interview systems emerged as a practical alternative to static question banks. These systems employed advanced language models to dynamically generate context-aware and role-specific interview questions, enabling interview difficulty and topic flow to adapt based on user responses [2]. This evolution significantly improved personalization and reduced operational costs, making interview preparation more accessible. Despite these improvements, many existing platforms still focus primarily on verbal interaction and provide limited analytical insight into long-term user progress.

Further research explored deeper response evaluation by analyzing how candidates structure and explain their answers during interviews. By examining coherence, reasoning depth, and content organization, these approaches provided more meaningful feedback compared to simple correctness-based scoring [3]. Although effective, many such systems were designed primarily for offline analysis and lacked seamless integration with real-time interview simulations.

To enhance realism, recent work incorporated speech-based interaction into mock interview platforms. Speech-to-text (STT) technologies enable spoken responses to be transcribed and evaluated automatically, improving accessibility and engagement compared to text-only systems [5]. While these approaches demonstrate strong performance under controlled conditions, challenges related to background noise, pronunciation variability, and transcription inaccuracies remain significant considerations for practical deployment [8].

Some recent interview preparation platforms have also explored the use of computer vision techniques to observe non-verbal behavioral indicators during interview sessions. Visual cues such as face presence and gaze direction can provide supplementary insights into user engagement and presentation behavior. However, many existing implementations rely on complex behavioral inference models, which may introduce interpretability and fairness concerns.

In addition, several systems leverage large language models to improve semantic understanding and feedback generation. These models enhance response evaluation capabilities but raise concerns regarding transparency, fairness, and explainability when applied to automated assessment scenarios [7]. Furthermore, most existing solutions focus on isolated components such as question generation, response analysis, or speech processing, lacking unified frameworks that support continuous performance monitoring.

In contrast, the proposed PREVUE.AI system addresses these limitations by providing an integrated mock interview platform built using the MERN stack. The system combines adaptive interview simulation, speech-based response evaluation, MediaPipe-based non-verbal observation, and structured performance analytics within a single web-based framework. This design emphasizes scalability, usability, and interpretable evaluation while avoiding overly complex behavioral inference models, making it suitable for practical interview preparation scenarios.

III. PROPOSED SYSTEM

The proposed PREVUE.AI system is designed as a web-based mock interview platform that provides realistic interview simulations and structured performance evaluation. The system aims to overcome the limitations of traditional interview preparation methods by offering adaptive interview flow, automated response evaluation, and performance tracking within a unified framework. The overall design focuses on scalability, accessibility, and consistency in assessment.

The system conducts role-specific technical and HR interviews by dynamically generating interview questions based on the selected job role, interview mode (technical or HR), and difficulty level. Adaptive question generation ensures contextual continuity throughout the interview, allowing the system to modify question complexity based on user responses. This approach closely resembles real-world interview patterns and enhances user engagement compared to static question banks.

To improve realism, user responses are captured primarily through voice input. Spoken answers are converted into text using automated speech recognition, enabling structured evaluation of response relevance, clarity, and completeness. The transcribed responses form the basis for automated scoring and feedback generation, ensuring objective and repeatable assessment.

In addition to capturing interview responses, the system performs structured performance evaluation using both verbal and non-verbal metrics. The verbal evaluation component analyzes the transcribed responses and assigns scores across

three key dimensions: correctness, depth of explanation, and response structure. Each of these metrics is scored on a scale of 0 to 100, and an overall verbal performance score is computed from these values.

The non-verbal evaluation component analyzes basic behavioral indicators captured through MediaPipe-based facial landmark tracking. This module estimates three presentation-related metrics: confidence, stability, and eye contact. Confidence represents the overall behavioral engagement of the user and is estimated using visual indicators such as eye contact consistency, blink rate, and face stability during responses. Stability represents consistent face presence throughout the interview session, while eye contact estimates whether the user maintains proper visual engagement with the camera. Each of these metrics is also evaluated on a scale of 0 to 100.

The final interview performance score is calculated by combining the verbal and non-verbal evaluation metrics to generate an overall score out of 100. This structured scoring framework allows users to understand specific strengths and weaknesses while enabling consistent and interpretable interview performance assessment.

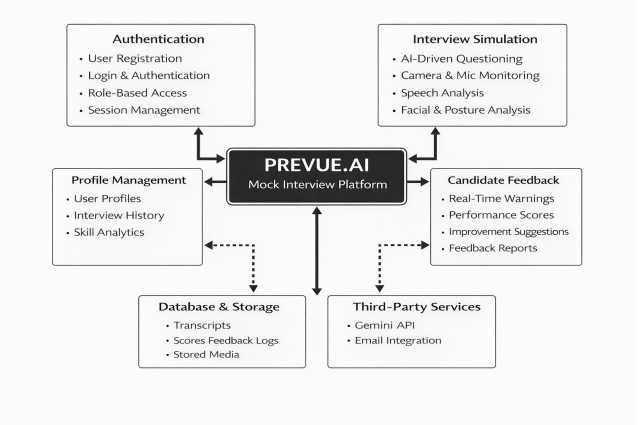

Fig. 1. System architecture of the proposed PREVUE.AI mock interview platform

Fig. 1 illustrates the overall system architecture of PREVUE.AI. The platform follows a centralized design where core interview functionalities are managed by the main system module, which coordinates user authentication, interview simulation, response processing, and feedback generation. Interview data and performance metrics are stored in a centralized database, while analytics and reports are generated to support continuous performance tracking. This architecture ensures smooth data flow between modules and supports scalability and real-time interaction.

All interview-related data, including scores, feedback summaries, and session details, are stored in a MongoDB database. A dashboard built using React.js and Node.js presents performance analytics, allowing users to review interview history, track progress, and identify improvement areas. By integrating adaptive interview simulation, speech-based evaluation, MediaPipe-driven visual observation, and performance analytics within a MERN-based architecture, the proposed system provides an effective and scalable solution for interview preparation.

IV. METHODOLOGY

The methodology of the proposed PREVUE.AI system follows a modular and web-based design aimed at simulating realistic interview environments while ensuring objective and scalable evaluation. The system workflow consists of adaptive interview generation, response capture, multimodal evaluation, and performance visualization. Each module operates independently while contributing to a unified interview preparation pipeline.

A. Adaptive Interview Generation

Adaptive interview question generation forms the core of the proposed system. Instead of relying on static question repositories, the system dynamically generates interview questions using large language models based on the selected job role, interview mode, and difficulty level. This context-aware approach allows the interview to evolve naturally by adjusting question difficulty and topic progression. Similar adaptive interview generation techniques have been shown to improve personalization and realism in automated interview platforms [2], [4].

B. Speech-Based Response Capture and Evaluation

User responses are primarily captured through voice input to replicate real interview conditions. Automated speech recognition is employed to convert spoken responses into textual form, enabling structured evaluation of answer relevance, clarity, and completeness. Prior research indicates that speech-to-text (STT) systems effectively support automated interview evaluation, although performance may vary with environmental noise and pronunciation differences [5]. The transcribed responses serve as the foundation for automated scoring and feedback generation.

C. Non-Verbal Observation Using MediaPipe

To complement verbal evaluation, the system incorporates basic non-verbal observation using MediaPipe-based facial landmark tracking. This module monitors visual indicators such as face presence, eye contact, and facial activity during interview sessions. The non-verbal observation module generates three presentation-related metrics: confidence, stability, and eye contact. These metrics are derived from facial landmark tracking patterns captured throughout the interview session. Confidence represents the overall behavioral engagement of the user and is estimated using indicators such as eye contact consistency, blink rate, and face stability during responses. Stability reflects consistent face presence throughout the interview session, while eye contact indicates visual engagement with the camera. Each metric is scored on a scale of 0 to 100. Non-verbal cues are treated as supplementary indicators of user attentiveness and presentation quality and do not involve emotion or sentiment detection.

Limiting visual analysis to essential facial landmarks reduces computational complexity and improves interpretability compared to emotion-based behavioral models [6].

D. Performance Analytics and Visualization

All evaluation results, including response scores, interview metadata, and performance summaries, are stored in a MongoDB database. The analytics module aggregates this data and presents it through an interactive dashboard developed using React.js and Node.js. This dashboard enables users to review interview history, compare performance across sessions, and track progress over time. Structured analytics and longitudinal performance tracking have been identified as key components for effective interview preparation systems [9], [10]. This modular methodology ensures scalability and practical usability of the proposed system. Additionally, the system generates a downloadable interview performance report for each completed session, summarizing verbal and non-verbal evaluation metrics, per-question analysis, and recommended improvement insights.

V. RESULTS AND DISCUSSION

The proposed PREVUE.AI system was evaluated through a series of controlled mock interview sessions to analyze its effectiveness in simulating realistic interview environments and providing structured performance evaluation. The testing process involved a group of undergraduate students who participated in approximately 25 mock interview sessions, including both technical and HR interview simulations using the platform. Each session consisted of dynamically generated interview questions, voice-based responses captured through the microphone, and real-time non-verbal observation through webcam-based facial landmark tracking. The evaluation focused on assessing the reliability of speech transcription, the stability of MediaPipe-based non-verbal tracking, the effectiveness of adaptive question generation, and the usefulness of the performance analytics dashboard. These experiments were designed to replicate typical interview scenarios and evaluate how effectively the system supports practical interview preparation.

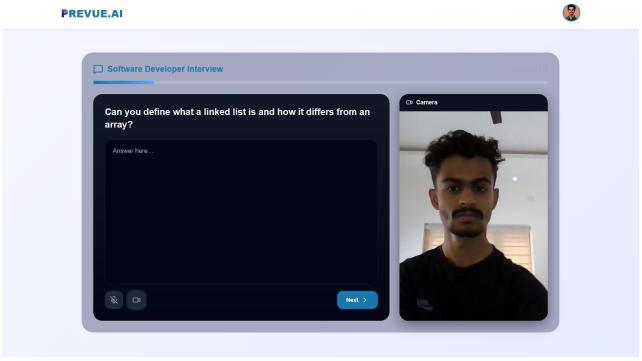

Fig. 2. User interface of the PREVUE.AI mock interview session

Fig. 2 shows the interview interface of the PREVUE.AI platform during an active mock interview session. The interface presents dynamically generated interview questions while capturing the user's spoken responses through the microphone. The spoken responses are automatically transcribed and displayed in the response textbox, allowing users to review the captured answer. The textbox can also be used to manually enter responses in situations where the answer is more suitable for typing rather than speaking. At the same time, the webcam feed enables real-time facial landmark tracking for non-verbal observation. This integrated interface allows the system to simulate a realistic interview environment while simultaneously collecting data required for verbal and non-verbal performance evaluation.

To evaluate the effectiveness of the proposed PREVUE.AI system, multiple mock interview sessions were conducted with undergraduate students preparing for technical and HR interviews. During each session, the system dynamically generated interview questions and recorded spoken responses while simultaneously monitoring non-verbal indicators through webcam-based facial landmark tracking. After completing an interview, users were able to download a detailed feedback report containing the evaluation scores for each question, including metrics such as correctness, explanation depth, and response structure. In addition, the platform dashboard enabled users to track their overall performance across multiple interview sessions, providing insights based on question difficulty levels and non-verbal evaluation metrics such as confidence, stability, and eye contact [9], [10]. The generated interview reports were also used to analyze individual question performance and identify common strengths and weaknesses across participants, enabling a more detailed evaluation of the system's scoring and feedback mechanisms.

The collected data from these interview sessions allowed the system to generate structured performance analytics and evaluate the reliability of its speech processing and visual observation components. Based on the experimental testing, the platform demonstrated stable operation during interview sessions while providing consistent scoring and performance tracking for users preparing for interviews. The key quantitative results obtained from the evaluation are summarized in Table I.

TABLE I
PERFORMANCE EVALUATION OF PREVUE.AI MODULES

System Module	Measured Accuracy (%)
Adaptive Interview Generation	88%
Speech-Based Response Capture and Evaluation	92%
Non-Verbal Observation Using MediaPipe	91%
Performance Analytics and Visualization	96%

The verbal scoring results were analyzed across different interview difficulty levels. It was observed that responses to easier questions generally achieved higher scores compared to medium and hard questions. During the evaluation sessions, the average verbal score for easy questions was approximately

91%, while medium difficulty questions obtained an average score of about 87%. For hard difficulty questions, the average verbal score decreased to around 82%, reflecting the increased complexity and depth expected in responses. This variation demonstrates the system's ability to differentiate response quality across different interview difficulty levels.

The MediaPipe-based non-verbal observation module operated reliably throughout the interview sessions. Facial presence was consistently detected during approximately 95% of the interview duration. Based on the collected observations, the average scores for the behavioral metrics were 90% for eye contact, 88% for confidence, and 92% for face stability. These results indicate that most participants maintained adequate visual engagement with the camera while responding to questions. Since the system avoids emotion and sentiment detection, the visual analysis remains computationally efficient and interpretable [6].

TABLE II
COMPARISON OF PREVUE.AI WITH EXISTING AI MOCK INTERVIEW SYSTEMS

Capability	**Existing Systems**	**PREVUE.AI**
Adaptive Interview Flow	Mostly static question flow	Context-aware questioning
Difficulty-Aware Evaluation	Usually not modeled explicitly	Evaluation across difficulty levels
Non-Verbal Behaviour Monitoring	Rarely used in interview platforms	Visual engagement tracking
Automated Interview Feedback Reports	Limited summaries or basic feedback	Detailed report for each interview

To further evaluate the practical capabilities of the proposed system, a comparison was conducted between PREVUE.AI and commonly available AI-based mock interview platforms. As shown in Table II, many existing systems support features such as automated question generation or speech-based interaction, but they often provide only basic analytics and limited support for continuous performance tracking. In contrast, PREVUE.AI integrates adaptive interview flow, speech-based response capture, visual engagement monitoring, and detailed performance analytics within a unified platform. Additionally, the system provides personalized performance tracking and downloadable feedback reports for each interview session, enabling users to monitor their progress and identify areas for improvement over multiple practice sessions.

Overall, the experimental results indicate that PREVUE.AI effectively combines adaptive interview simulation, speech-based evaluation, basic non-verbal observation, and structured analytics within a unified web-based platform. While certain limitations related to speech recognition accuracy under noisy conditions remain, the system achieves its primary objective of providing a scalable, realistic, and objective interview preparation environment suitable for real-world use.

VI. CONCLUSION AND FUTURE WORK

This paper presented PREVUE.AI, an AI-driven mock interview platform designed to provide realistic interview simulations and structured performance evaluation using a web-based architecture. By integrating adaptive interview question generation, speech-based response evaluation, MediaPipe-based non-verbal observation, and performance analytics within a MERN stack framework, the proposed system addresses key limitations of traditional interview preparation methods. The system enables consistent assessment, reduces dependency on human interviewers, and allows users to track performance improvement across multiple interview sessions. Experimental evaluation demonstrated that PREVUE.AI effectively simulates real interview conditions while maintaining scalability, interpretability, and usability.

Although the proposed system achieves its primary objectives, several enhancements can be explored in future work. Non-verbal analysis can be extended to include posture stability and controlled gesture tracking to enrich presentation-related feedback while maintaining transparency. Speech evaluation can be improved by incorporating advanced fluency metrics such as pause duration and speaking rate analysis. Future versions of the system may also support multilingual interview simulations to improve accessibility for a broader user base. Additionally, role-specific evaluation rubrics and interviewer-style customization can further enhance realism. Expanding the platform with institutional dashboards for educators or recruiters would enable large-scale training and analytics. These enhancements would further strengthen the adaptability, accuracy, and applicability of PREVUE.AI in diverse interview preparation scenarios.

REFERENCES

[1] S. Senthilkumar, R. Prakash, and K. Suresh, "AI-Based Mock Interview System Using Natural Language Processing," *International Journal of Engineering Research & Technology*, vol. 12, no. 4, pp. 520–525, 2023.

[2] A. Mishra, P. Verma, and S. Gupta, "AI-Driven Virtual Mock Interview Development," *Proceedings of the International Conference on Artificial Intelligence and Applications*, IEEE, pp. 112–118, 2024.

[3] Y. Chou, L. Wang, and H. Chen, "An AI Mock Interview Platform for Interview Performance Analysis," *Journal of Intelligent Systems*, vol. 32, no. 2, pp. 145–156, 2023.

[4] R. Sharma and N. Patel, "Automated Interview Question Generation Using Transformer Models," *International Journal of Computer Applications*, vol. 185, no. 6, pp. 12–18, 2022.

[5] D. Yu and L. Deng, "Automatic Speech Recognition: A Deep Learning Approach," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 18–31, Jul. 2018.

[6] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1867–1874, 2014.

[7] G. ShivajiRao et al., "AI-Powered Virtual Job Interview Simulator Using Natural Language Processing," in *Proceedings of the 8th International Conference on Trends in Electronics and Informatics (ICOEI)*, IEEE, 2025, doi: 10.1109/ICOEI59956.2025.13824299.

[8] X. Li, J. Yao, and Z. Wang, "Robust Speech Recognition in Noisy Environments," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 2105–2116, 2021.

[9] M. Hiremath and S. Kulkarni, "Performance Analytics in AI-Based Recruitment Systems," *International Journal of Data Science and Analytics*, vol. 11, no. 3, pp. 221–230, 2022.

[10] T. Brown et al., "Data-Driven Evaluation Techniques for Automated Interview Systems," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 1, pp. 1–22, 2023.