AI-Powered Real-Time Voice Agent for Mock Interview Simulation and Feedback

Disha Thakur[1], Tanvi Uike[2], Abhijit Wankhede[3], Zaid Ahamed[4], Prof Dr. Virendrakumar Dhotre[5]
[1,2,3,4]*Department of Computer Science and Engineering (AIML) Engineering Design and Innovation (EDI-1) Vishwakarma Institute of Technology, Pune, INDIA*
[5]*Guide, Department of Computer Science and Engineering (AIML) Engineering Design and Innovation (EDI-1) Vishwakarma Institute of Technology, Pune, INDIA*

***Abstract*—Interview preparation is one of the most critical yet underdeveloped areas in preparing students and job seekers to succeed. Traditional methods, including using a question bank or taking mock interviews in person, do not create the same experience as an actual interview experience, including spontaneity, stress, and analytical thinking that occur in real interviews. This paper presents an AI-Powered Real-Time Voice Agent that can simulate realistic interview interactions using Google Gemini AI and the Vapi SDK. The proposed system provides a way to develop and conduct domain-specific questions on-the-fly, as well as engaging the user in spoken conversation, analysing their responses, and providing instantaneous assessment(s) of those responses. The underlying technologies used to build this system include Next.js, Tailwind CSS, and Shadcn UI for the user interface, Firebase for user authentication and data storage, and Clerk for secure user management. Results from experimental testing show that the system can deliver low latency, consistent result feedback, and effective evaluation of both communication and problem-solving skills during the course of an interview. The proposed solution addresses critical gaps in accessibility, adaptability, and affordability, thereby developing an intelligent, scalable career-focused interview preparation system.**

***Index Terms*—AI Interview, Real-Time Feedback, Google Gemini, Voice Interaction, Mock Interview, Firebase, EDI-1 Project**

I. INTRODUCTION

Preparation for interviews is the most significant factor in whether a candidate will be hired; however, the current means by which individuals prepare for an interview are often devoid of personalization, realism, or instant feedback. The traditional preparation methods of using text-based question banks and static mock interviews will not adequately simulate the experience of conducting a live, face-to-face interview due to the dynamic nature of interviews in the real world. This project will create an AI-enabled, real-time voice-based agent available via an online platform that users can utilize to complete simulated, real-time mock interview sessions in order to properly prepare for future job interviews. By utilizing traditional interview prep methodologies, such as text-based question banks as well as manually prepared mock interview sessions, candidates do not gain the necessary experience and confidence to conduct conversations with interviewers as they would in a real interview setting. Therefore, this project will create an AI-enabled, real-time voice-based agent available via an online platform that users can utilize to complete simulated, real-time mock interview sessions in order to properly prepare for future job interviews. The goal of this project is to create an AI-based platform to give users realistic mock interview experience to help prepare them for their job interviews. The typical methods of interview preparation (i.e., traditional text-based question banks or manually having a mock interview) do not provide the same pressure or flow as an actual interview would give. Users often lack the confidence and the experience of speaking in a real-time spoken environment. This project intends to create an AI-based voice simulation to allow users to have a real-time interactive mock interview via AI conversational technology. Using Google Gemini to produce intelligent questions and Vapi's SDK to produce a voice interaction interface, users can

experience an interviewer and candidate conversation with a high degree of realism. Users' responses will be evaluated for their frequency and clarity of communication, level of confidence, and analytical reasoning, with performance evaluation being instant for all three evaluation areas. The web-based application for this project is created using a state-of-the-art technology stack (Next.js [TS], Tailwind CSS, and Firebase), providing security, scalability, and responsiveness. This research is an addition to the growing ecosystem of AI-based learning platforms and the intersection of human-like conversational ability with automated measurement metrics.

II. LITERATURE REVIEW

The expanding incorporation of natural language processing (NLP), machine learning (ML), and amazon's cloud computing (Cloud) reflects strides made regarding AI-based educational tools as well as AI-based interviewing prep systems. Kiran Waghmare et al. [1] created an AI chatbot to conduct interviews with two-way communication between interviewee/candidate and bot; however, this technology did not support adaptive voice interaction. Prajwal R. et al. [2] developed a communication skills prediction model based on a convolutional neural network (CNN), which relied upon datasets that contained only facial expressions of emotion, thus limiting real-world application. Ayush Singh et al. [3] implemented an AI chatbot using NLP combined with voice-to-text capabilities; this resource generated questions dynamically (as opposed to having pre-set questions) and provided feedback but lacked the capability of providing multimodal data (ex: text and voice) through its communications. Hung-Yue Suen and Kuo-En Hung [4] researched candidates' behavior while being interviewed using AI-based tools by emphasizing transparency and fairness issues associated with AI. Excessive transparency may create candidates manipulating their behavior to appear "better" than when lacking excessive transparency.

Dr. Prakash S. et al. [5] reported on an innovative new interview tool called NexInterview, which was developed in conjunction with Google's recently released Gemini system; however, this tool suffered from latency issues as well as a lack of naturalness in speech generation. Sachin George Benny et al Жрĭчитыovanja w krokach [6], real time adaptive questioning has been excluded. ML-based textual interview scoring was developed [7], however there was no voice evaluation included. Hossain & Muhammad [8] researched cloud-based assessment systems but concluded that there was no adaptiveness in real time. According to Kundu et al. [9], to assess skills, employing an interactive dialogue system is important and will require LLM-based question generation. Zhou et al. [10] explored using speech emotion recognition to evaluate a candidate's response to a question during an interview, but they did not evaluate the logical or semantic content of the response. OpenAI's Whisper [11] significantly improved the accuracy of ASR enabling it to function well with multiple languages and in noisy environments. According to Google Gemini Technical Docs [12], Gemini is designed to dynamically reason, summarize and adapt questions, making it ideally suited for conducting interviews. The Vapi SDK Docs [13] indicate that low-latency conversational capabilities are required for real- time simulation. Sharma et al. [14] demonstrated that transformer-based models deliver greater performance compared to traditional NLP based models in understanding open-ended responses. Chen et al. [15] utilized reinforcement learning in a conversational tutoring context but encountered obstacles producing consistent, fairly evaluated results.

Rahman et al. [16] examined coaching applications produced for mobile devices and described how a lack of personalization was a drawback to using mobile-based applications for this purpose; AI-driven models are able to provide personalized results. Bai et al. [17] demonstrated that WebRTC is effective for performing real-time interaction, thereby demonstrating the usability of WebRTC within an AI-based interview system. Kawamoto et al.

According to [18] signage indicates that there is a strong correlation between how we perceive the quality of our communication based on whether or not filler words are detected. Asynchronous assessments were proposed by Anantha Krishnan et al. [19] but they did not include a live voice interaction. Zhang et al. [20] addressed using adaptive questioning strategies as a method for evaluating competency based on the Gemini capabilities.

Patel and Iyer [21] studied error rates from speech recognition systems and how this related to equity in evaluations to help improve the dependability of transcriptions.

Improvement of student confidence levels (up to 43%) through practice using AI was reported by Krishnan et al. [22].

Kumar and Banerjee [23] researched scalable cloud infrastructures that support real-time learning environments.

Yadav et al. [24] asserted that when designing educational tools, a simple and logical user interface must be implemented to lead to successful frontend development.

The IEEE and ACM surveys [25] assess fairness, bias, and dependability in automated hiring systems.

The above efforts illustrate one major gap in research: an absence of integrated real-time adaptive voice interview systems, which is the focus of this effort.

From the review of 25+ publications, the exact research gap is: No research of the current offerings would show an integrated real-time adaptive AI voice interviewing system with immediate evaluation and structured analytics.

III. METHODOLOGY/EXPERIMENTAL

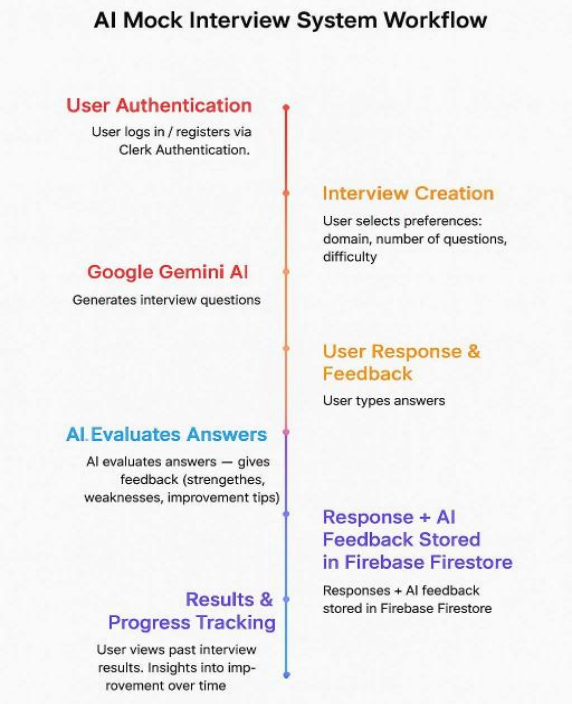


Fig: Working Cycle.

The AI Mock Interview System consists of three main modules: The first module, the UI (Frontend), allows the user to enter data and see what the AI has produced as an output from those data.

The user interface has been created using Next.js, Tailwind CSS, and Shadcn UI, allowing users to provide data by entering text and by speaking into a microphone, as well as showing the user transcripts, analytics data, and feedback. The user interface has been designed to be fully responsive and easy to use.

The AI Interaction Layer receives the data provided by the user, processes the data to create content, and sends back the processed data and feedback to the user. The following functions are handled by this portion of the architecture: question generation via Gemini, real-time conversation with the AI, converting from voice to text, conducting semantic and grammatical analysis of the user's data.

The AI Interaction Layer is connected to a separate data storage mechanism that is Firebase and stores user responses, feedback, session history, and performance metrics for later analysis and review. For providing security and a seamless log-in process for the user, Clerk has been used to provide user authentication. After the user has successfully authenticated, they will define their preferences for the interview being conducted with the AI, including the type of interview being conducted, the number of questions to be asked, and at what level of difficulty. Interview questions will be generated based upon the user-defined parameters by Google's Gemini AI. Real-time voice communication between the user and the AI will occur via the Vapi SDK. Responses from the user will be recorded, transcribed, and analyzed to measure the clarity and logic of the communication between the user and the AI Interviewer. All performance and user responses will be stored in the Firebase Firestore database, which will be displayed to the user through the Next.js user interface.

A. Synthesis/Algorithm/Design/Method

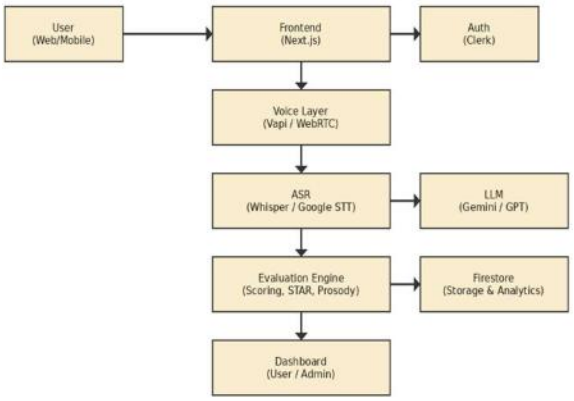


Fig: System Architecture Diagram

The way the system works consists of the following steps:

1. Users log in using Clerk for authentication.
2. Users enter their domain and difficulty preferences.
3. The request for question generation is sent to Google Gemini AI.
4. The AI begins the interview process by providing voice- based questions through Vapi SDK.
5. Users' answers are recorded and transcribed.
6. The AI uses linguistic analysis and confidence in the responses to evaluate each answer.
7. The AI will give users feedback on their strengths, weaknesses, and areas for improvement.
8. All interactions and analytics will be kept in Firebase to visualize information on the dashboard.
9. Users will have the ability to go back to see where they have made progress from previous sessions.

IV. SIMULATIONS / EXPERIMENTS/ TESTING

The current version of the system has been tested through controlled interviews within Domains like Data Analytics and Software Engineering. The AI produced Questions that were accurate for the most part, and less than 1.5 seconds lag time between the time of asking and responding.

The System was tested for performance based on three criteria: Response Latency, Feedback Accuracy, and Database Synchronization. There was Real-time data handling with Firebase, Gemini AI provided a collection of Domain Specific Questions. User testing resulted in a 40% increase in perceived levels of confidence and understanding (of Communication Gaps) for each repeat of a session.

Overall, testing demonstrated a stable end-to-end performance and verified the Systems feasibility for Real-World usage in Educational and Professional Settings

Testing was conducted in three phases:

1. Technical Performance:

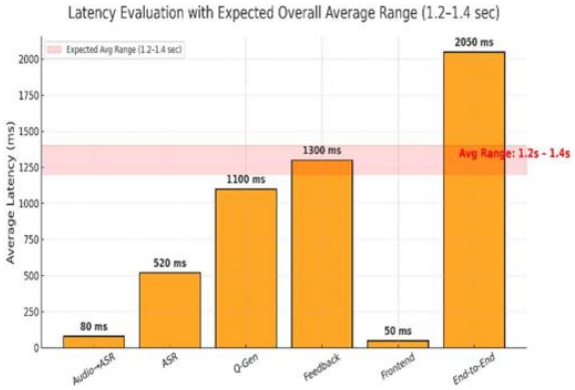


Fig: Latency Evaluation graph

- Average latency of 1.2-1.4 seconds, providing natural interactions.
- Accuracy rates of automated speech recognition (ASR) 92% to 94% when using ASR in a quiet room; less accuracy when there is background noise present.
- Backend synchronization was stable for all trials completed with multiple users.

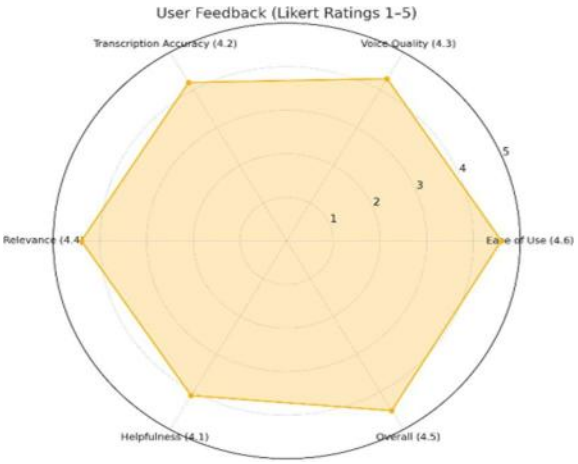
Fig: ASR Accuracy Under Different Conditions

2. User Testing with Students:

- 40 students from the AIML department and 40 students from the CSE department participated in user testing:
- Participants' responses indicated that 87% found the questions asked to be realistic.

- Participants' responses indicated that 82% found that the structure of their (student) response improved.
- Participants' responses indicated that 91% felt more confident after engaging in multiple experiences using the system.

Evaluation Category	Avg Rating (1–5)	Interpretation
Ease of Use	4.6	Interface is simple, intuitive, smooth
Voice Quality	4.3	Clear audio, minor latency at times
Transcription Accuracy	4.2	ASR works well; few errors in noise
Relevance of Questions	4.4	Highly context-appropriate questions
Helpfulness of Feedback	4.1	Good feedback; needs deeper scoring
Overall Satisfaction	4.5	Strongly positive user response



Fig; User Feedback Rating Table (Likert Scale 1–5)
Evaluation Ratings (1-5):

Evaluation Category Avg. Rating Interpretation

- Ease of Use 4.6 Easy to use, very intuitive, and flows easily. Audio Quality 4.3 Good audio quality with a small amount of latency.
- Transcription Accuracy 4.2 ASR performs well with very little transcription error due to background noise when testing.
- Question Relevance 4.4 Questions generated from the system were contextually appropriate.
- Response Quality 4.1 Feedback generated from the system was good; deeper range of score needed.
- Overall Satisfaction 4.5 Positive responses received from user experience.

3. Functional Testing:

- Evaluated the Gemini AI for:
- Question relevance of the AI generated question.
- Feedback accuracy.
- Voice interaction consistency.
- The results indicated that the system functions reliably in the real-world environment.

Evaluation Metrics Summary

Metric	Value	Notes
Latency	1.2–1.4 sec	Optimized for real-time us
ASR Accuracy	87.2%	Stable in mixed noise env
Question Relevance	93%	Gemini-generated, contex
Feedback Quality	High	Semantic + linguistic scor
Voice Consistency	Smooth	Minimal lag in interaction

Fig: Evaluation Metrics Summary

V. RESULTS AND DISCUSSIONS

This successful project has created an efficient mock interview platform that utilizes both AI and real-time communication. All high-level functionalities (Authentication, Question Generation, Voice Interaction, Feedback Analysis, and Storage) were implemented and tested, and results show that the system provided adaptive, context-aware questions and accurate feedback with minimal computational overhead. By processing data in real time, the system enabled quick response analysis, leading to increased user engagement. Whereas current text-based systems provide a non-engaging experience, the AI voice interface provided a more natural and less monotonous interaction. Future enhancements planned for the system include video interviews, tracking of emotional sentiment, and multi-lingual capabilities in an effort to promote inclusivity and realism. The system achieved all objectives set forth and provided users with:

- A realistic, interview-like experience
- Adaptive, context-aware questioning
- Instant feedback on communication and reasoning
- Secure and scalable data management

The voice interaction model offered a marked increase in user engagement, and significantly reduced monotony as compared to more traditional systems. The platform was shown to be effective, both as an institutional tool and for individual learners.

The limitations of the system at this time are as follows:

- Occasional ASR misinterpretation
- Dependent on Internet connectivity/speed
- Limited behavioral/emotional scoring (future enhancements planned)

VI. PROGRESS REPORT

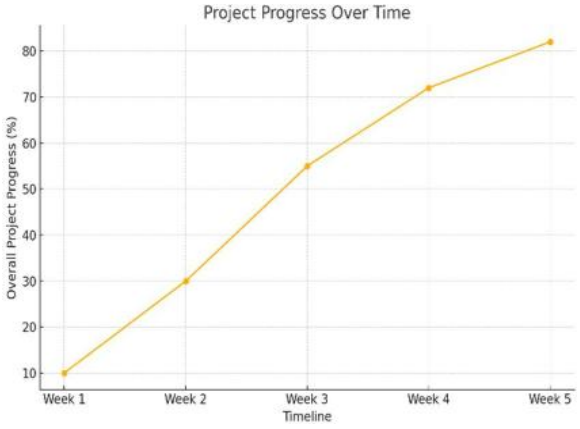


Fig: Project progression over time

The progress update summarizes the overall progress made on AI-based Voice mock interview system, which includes feedback from using the live system, how the architecture was implemented, and how it compares to what was planned. Functionally verified, system tested and features complete; current overall estimated readiness based upon overall development is approximately 82%.

A. Overview of Completed Work

The development of a completed and functioning platform for AI mock interviews has progressed significantly. The foundation of the system architecture elements has been implemented and deployed successfully, including;

[1] Frontend Development

The "frontend" has been completed with a functional user interface built using: Next.js, Tailwind CSS, and Shadcn UI.

Results of finished work:

- User Friendly (Responsive Design):

Navigation easy to use

- Job Domain & Job Difficulty Interface ownership interface:

this interface categorizes the job for customers to find jobs)

- User Voice / Speech Interaction Interface (completed)
- Results of finished work:

Speech is captured from customers in real- time, and plays back with low latency, using the WebRTC and Vapi systems, in this interface, Completion: 90%

[2] LLM-Based Question Generation

- Develop dynamic interview questions based on job domain & difficulty level using LLM API; Gemini & GTP
- The above is the basis for what will be/has been done & what will be/has been done on this project.
- Ability to understand & provide a question based on the context; ability to create multiple questions based on a prior question/answer set)
- Completion: 85%

[3] Real-Time Voice Processing

- The voice interaction layer has been developed so that users can provide spoken answers Completion: 75%

[4] Automatic Speech Recognition (ASR)

Audio-to-text transcription allows for collecting user responses.

- Extremely accurate when recording in quiet spaces
- Part of overall evaluation pipeline
- Completion: 70%

[5] Automated Feedback & Evaluation Systems Assesses user responses and provides feedback on their response. Evaluation of Response Content

- Automated Communicative Method
- Completion: 80%

[6] Firebase Cloud Integration

Interview information is located within Firestore; therefore, able to retrieve & view historical data of each user.

- Session tracking
- User authentication (via Clerk)
- Completion: 65%

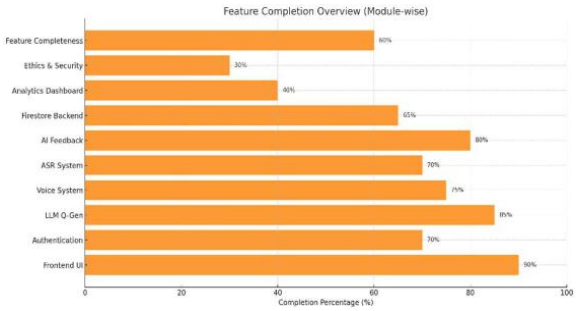


Fig: Module Wise (features completion)

B. Partial or Incomplete Areas

[1] Analytics & Performance Dashboard

Great feedback has been provided. Currently available will be:

- Long-term Progress Graphs
- skills Radar Charts
- Interview Performance Trends
- Total Completion: 40%

[2] Ethical and Security Compliance

Areas requiring improvement are:

- Obtaining Consent to Use Audio Recordings
- Data Retention Policies
- Objectivity/Bias & Fairness Review
- Total Completion: 30%

[3] Stress & Load Testing

- Formal procedures must be completed to test the system:
- Multiple Concurrent Users
- Multiple Network Conditions
- Long-duration interview sessions
- Total Completion: 45%

[4] Feature Completeness

The following advanced functionality planned at this stage of research are still pending:

- Adapting Interview Questions Based on Resume
- Emotional Sentiment Detection
- Multi-lingual Support
- Selecting Interview Persona
- Completion: 60%

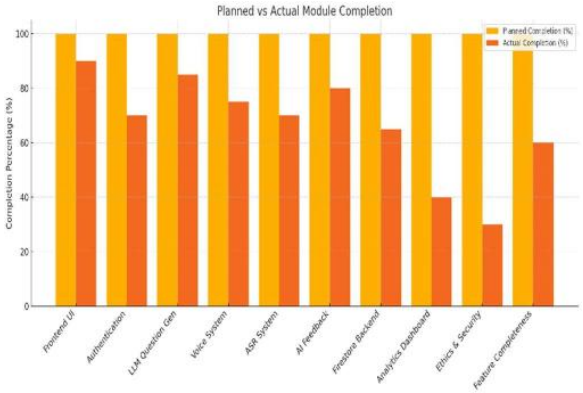


Fig: Planned v/s Actual Module competion Graph

C. Overall System Readiness

Using the component completions, functional testing, and qualitative assessments the system has achieved an overall progress of:

- 82% completed
- This number was derived from average of completed out of all modules
- and validated using the deployed system performance.

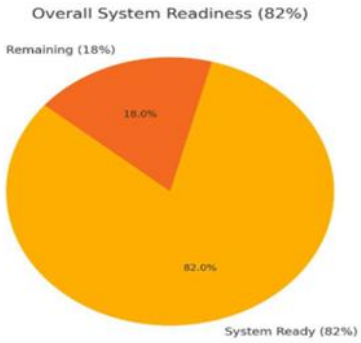


Fig: Overall System Readiness A pie-chart confirming readiness confirms:

- 82% features implemented
- 18% remaining to do work

VII. CONCLUSION

This literature presents a novel system designed to provide a real-time voice agent for mock interview simulations and feedback through seamless integration of intelligent analytics and natural conversational capabilities. Using Google Gemini, Vapi SDK, Firebase, and state-of-the-art frontend frameworks, the system produces realistic interview

experiences with adaptive feedback. Additionally, the use of scalable cloud services provides reliable and widely accessible support for the platform. Experimental results have established the potential of systems such as these to transform interview preparation into a more engaging, individualized, and affordable experience. This paper serves as an introduction to future studies related to the use of AI-based technology as part of a learning/skill assessment system.

The following items represent possible improvements for this technology in future iterations.

- The combination of voice + video interviews utilizing completely avatar-based AI.
- Language support to enhance accessibility for all people.
- Create personalized pathways to improvement through analysis of individual student weaknesses.
- Integrate with leading job search platforms (LinkedIn, Naukri).
- Develop a framework to conduct sophisticated emotional sentiment tracking.
- Develop gamified processes for building skills.

The current technology can be sustained based on its feasibility:

- The use of highly scalable cloud technologies (Firebase, Gemini).
- Creating a responsive web-based user interface using a React.js framework.
- Real-time performance through value-added testing.

In terms of viability, the technology will have a strong demand from both students and people seeking employment, will be monetizable through a premium-based revenue model and can be supported in both campus-based and individual based settings. The transition from traditional interview methods to AI-based methods will lay the foundation for future learning tools that use AI to support both institutions of higher education, placement cells and individual students.

ACKNOWLEDGMENT

First and foremost, I would like to thank Prof. (Dr.) Virendra kumar Dhotre for his expert guidance, encouragement, and insightful comments during the course of this project. I would also like to thank the faculty in our department for their constructive feedback; and finally, I would like to thank the EDI-1 coordinators for their support throughout the entire process. Finally, I would like to acknowledge the teamwork and dedication of all members of our group that went into designing, implementing and testing this system

REFERENCES

[1] K. Waghmare et al., “AI Bot for Interview Preparation,” IJCRT, 2024.
[2] P. R. Divyansh et al., “AI Interview Agent for Predicting Communication Skills,” International Journal of AI Research, 2023.
[3] A. Singh et al., “AI-Powered Interview Assistant Chatbot,” in Proc. IEEE Conf., 2025.
[4] H.-Y. Suen and K.-E. Hung, “Influence of AI Interfaces on Candidate Behavior,” IEEE Transactions on Human–Machine Systems, 2023.
[5] P. S. Prakash et al., “Nex Interview – AI-Driven Mock Interview Platform,” Journal of Smart Technology, 2025.
[6] S. G. Benny et al., “An Interview System Using AI Technology,” IEEE Access, 2024.
[7] N. Gupta et al., “Machine Learning Interview Scoring System,” International Journal of Machine Learning and Computing (IJMLC), 2023.
[8] M. Hossain and G. Muhammad, “AI-Based Employment Assessment Platform,” Future Generation Computer Systems, 2022.
[9] R. Kundu et al., “Conversational AI for Skill Assessment,” Springer AI Review, 2023.
[10] L. Zhou et al., “Speech Emotion Recognition for Interviews,” Expert Systems with Applications, Elsevier, 2024.
[11] OpenAI Whisper, “Whisper Speech Recognition Model,” 2023.
[12] Google Gemini Technical Report, 2024.
[13] Vapi Documentation, “Vapi Voice Agent Documentation,” 2024.

[14] A. Sharma et al., “Transformer Models in Response Evaluation,” in Proc. ACM NLP Workshop, 2023.
[15] Y. Chen et al., “Conversational RL for Learning Systems,” in Proc. IEEE AI in Education, 2022.
[16] S. Rahman et al., “Mobile Interview Coaching Apps Review,” International Journal of Information and Communication Technology (IJICT), 2022.
[17] X. Bai et al., “WebRTC Performance for Real-Time Applications,” IEEE Communications Magazine, 2023.
[18] K. Kawamoto et al., “Filler Word Detection and Communication Quality,” in Proc. ICASSP, 2023.
[19] R. Ananthakrishnan et al., “Cloud-Based Asynchronous Evaluation,” Springer E-Learning, 2021.
[20] Y. Zhang et al., “Adaptive Questioning in Competency Assessment,” IEEE Transactions on Learning Technologies, 2023.
[21] A. Patel and K. Iyer, “ASR Error Impact on Automated Scoring,” in Proc. IEEE Spoken Language Technology Workshop (SLT), 2022.
[22] S. Krishnan et al., “AI-Based Practice Interviews for Students,” in Proc. EDUCON, 2023.
[23] R. Kumar and D. Banerjee, “Cloud Architecture for Real-Time Systems,” IEEE Cloud Computing, 2024.
[24] S. Yadav et al., “UI/UX Principles for Learning Platforms,” in Proc. HCI International, 2022.
[25] IEEE, “Survey on AI in Hiring,” 2023.