

A Real-Time AI Voice Agent for Interactive Interviews

K B Bini¹, Bhargav M², Likitha D G³, Madhumitha T S⁴ and Mohith Prakash⁵

¹Department of Artificial Intelligence and Machine learning, AMC Engineering College, Visvesvaraya Technological University, India

Email: binikb.shr@gmail.com

²⁻⁵AMC Engineering College, Department of Artificial Intelligence and Machine Learning, Visvesvaraya Technological University, India

Email: gowdabhargav3@gmail.com, likithadg604@gmail.com, madhumithats1708@gmail.com, mohithprakash26@gmail.com

Abstract— This system introduces an AI-powered-interview platform that helps authentic job interview experiences through hands-on practice. The system integrates techniques of Natural Language Processing (NLP), Speech to Text conversion and vice versa. In addition to this Large Language Models (LLMs) like Gemini is used to evaluate the overall skills, basically the technical and communication of candidates dynamically. Using this model candidates can experience different versions of adaptive questioning constrained with time. Real-time interaction through browser is achieved through WebRTC and Vapi SDK. Whisper is the Open Source speech recognition model turning speech into text, voice engines making replies that sounds like a human. This framework also provides personalized feedback to the candidates through an assessment module large language models. Options are given to candidates for selecting the level of complexity of questions, clear responses at will, expandable mock interviews towards enrichment of skills.

Index Terms— Natural Language processing (NLP), Speech-to-text, mock interviews, Feedback.

I. INTRODUCTION

Upgrading with proper technical skills and making yourself prepared for interview seems to be the most challenging aspects of professional development. Existing methods of mock interviews are often constrained by the good human evaluators, a lack of scalability, and subjective feedback. With the advancement in technology, it is possible to simulate a realistic interview scenario that provide adaptive questioning, immediate assessment, and structured feedback to candidates.

This paper introduces a real-time AI-powered interview platform that integrates speech-to-text (STT), text-to-speech (TTS), large language models and Natural language processing (NLP) to replicate authentic interview experiences. Our system is developed with a multi-level framework, time-slots and retry mechanisms to enhance candidate performance. Real-time interaction is achieved through WebRTC and Vapi SDKs for seamless communication and structure data storage using SQL Databases, the system ensures scalability and performance. The main focus of this system lies in bridging gap between traditional interview process and AI-driven professional training. This platform not only evaluates soft skills and technical skills but opens a way for continuous improvement of the candidates through integrated LLM-based practice, making it worth for jobs seekers.

A. Motivation

The mock interview systems apply a pre-determined set of questions and are evaluated by humans, or use simple scoring. Hospitality is composed of several disadvantages in traditional methods. It would also not be easy to scale it to thousands of users. Another problem that will lead to a mismatched score in the performance of the candidates is the human judgment. Moreover, they lack real-time behaviour assessment of the systems. Some of such systems fail to apply multiple feature cues, including speech pattern, tone and response of adaptive feedback. The system proposed will overcome this challenge through the combination of the strength of natural language processing induced question generation, speech detection and constant feedback to ensure a compelling interactive assessment process.

II. RELATED WORKS

Conventional approaches as well as mock interviewing platform usually relies on the process of rule-based or filtering and restricts due to data sparsity and inability of real-time flexibility.

Sharma et al. (2023): AI-Powered Interview Platform with Computer Vision, NLP and Generative AI [1] created an AI-based mock interview based on computer vision, NLP and generative AI to assess the candidates. Generative AI made it possible to have realistic interviewer like interaction, enhancing interaction and quality of responses. The system examined non-verbal signals so as to produce feedback reports. Nevertheless, the system was not as adaptable in terms of real time since the feedback was provided upon the completion of the interview. Wongso et al. (2022): AI Mock-Interview Platform for Interview performance Analysis [2] designed a performance monitoring and training of interviews based on AI. The suggested system applies behavioural and linguistic analysis to check on the performance of the candidates and offers structural feedback to the interview; this increased the performance of the preparedness of students or real-time interviews. But the flexibility in real-time and deployment process was restricted in the research.

Jagtap et al: AI-Based Interview Agent [3] presented an interview agent based on AI which is an interviewer agent which utilizes natural language processing and computer vision to simulate a natural interview situation. The system marks a candidate according to the fluency and relevance of his or her response, and inspects non-verbal facial expressions. This has contributed to picking of more competent candidates through follow up of valuable reviews of candidate behavior, decreased human assessors as well as making the program more expansive. Their model, therefore, relies on the classic verbal reactions; it stumbles upon the additional context and adjusting to other fields. These points, which refer to the enhancement of the completeness of an interview, should be improved.

Pathak and Pandey: A Multi-agent System to Interview, evaluation and candidate scoring [4] suggested a multi-agent system to recruit employees using AI. On their part, they involve varying agents, including one agent to interview, one to evaluate and another to score; they have different jobs. That is how the architecture may be modular, scaleable, efficient and less biased. Since the agents operate in a distributed fashion, multiple candidates can be processed at the same time. Findings showed that the workload of recruiters decreased, and the large-scale hiring was more consistent. Nonetheless, there are still issues in the explainability, domain-specific assessment, and human supervision.

All these studies highlight the need to incorporate AI-based approaches to effective and scaleable interview simulation and candidate assessment. Our suggested system will be based on these ideas and will include live interaction, adaptive learning, and multi-modal support (text, voice, and video) thus providing candidates with personal feedback and an interview experience.

III. PROPOSED SOLUTION

The system integrates AI-driven question generation, real-time communication, adaptive scoring, and personalized feedback into a single framework. Its primary components include:

A. AI-Driven Question Generation

The Gemini (large-language model) generate domain-specific and context-sensitive interview questions dynamically. The questions evolve based on prior user responses, overall interview performance and ensuring interview experience.

B. Real-Time Voice Interaction Layer

WebRTC (Web Real-Time Communication) is an open-source framework that enable real-time audio, video and data exchange directly between browsers and combined with socket-based communication enable low-latency

bidirectional audio transmission between client and server. The system behave like a live interviewer, providing immediate responses to user input.

C. Speech and Language Processing

Whisper is a transformer-based sequence-to-sequence recognition model developed by OpenAI. It listens to the candidate's speech and converts it into text in json format using transformer-based model. This text is then passed to Gemini, which analyses the candidate's behaviour and answers using the rules and parameters we've defined for the Gemini to work.

D. Adaptive Performance Review

The scoring model evaluates users across multiple competencies, including:

- Enthusiasm & Interest
- Communication Quality
- Self-Awareness & Reflection

Scores influence the difficulty and nature of subsequent questions, creating an adaptive interview pathway.

E. Personalized Feedback Engine

Following each session, the system produces structured feedback summarizing user strengths, weaknesses, and recommended improvements. The platform also maintains historical data to support long-term skill development.

IV. SYSTEM ARCHITECTURE

A. The System Architecture Comprises Five Layers

Client Interface Layer

The browser-based interface built with Next.js and Tailwind CSS, enabling seamless interaction through audio and optional video.

Application Layer

Handles authentication, session management, and secure communication routing. Flask-SocketIO servers process real-time events.

Database Layer

User profiles, interview history, transcripts, and performance metrics are stored in Firebase based systems.

AI Processing Layer

Responsible for:

- Speech recognition (Whisper),
- Question generation (LLMs),
- Semantic evaluation,
- Sentiment and tone analysis.

Feedback and Analytics Layer

Scores, insights, and performance summaries are generated and presented through visual dashboards and analytical graphs.

B. Block Diagram Explanation

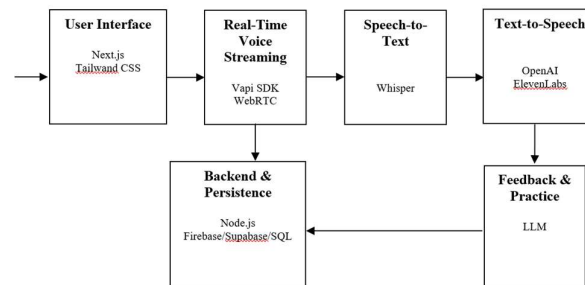


Figure 1. Block diagram

C. System Flow

User interaction begins when the candidate speaks directly into the interface, enabling real-time communication between the user and the system. The backend infrastructure, powered by Node.js, handles authentication, event processing, and server-side logic, while structured data—including transcripts, user profiles, and performance summaries—is securely stored in Firebase, Supabase, or SQL databases. Low-latency voice streaming is supported through Vapi and WebRTC, ensuring smooth audio transmission throughout the interview session.

The system language pipeline integrates several AI libraries to handle tasks. Whisper converts spoken input into text, neural Text-to-Text (TTS) systems engines such as OpenAI or ElevenLabs generate natural sounding responses, large-language models interpret user intent to produce coherent, content-aware replies. The front-end interface, built using Next.js and Tailwind CSS, provides an intuitive user experience, while the audio maintains synchronization between speech-to-text, text-to-speech and backend. This combination of pipeline enables seamless two-way voice interaction that closely resembles a real-time interview environment.

To maintain system efficiency, management and model storage are optimized for fast loading and minimal response delay. Supportive components such as feedback tools, memory systems and structured guidance modules enhance the user experience by personalized interactions based on past behaviour and performance. Comprehensive documentation using workflow, API routes and rules set ensures adaptability, scalability and smooth collaboration cross development.

Further, our system is designed with allowing different components such as STT modules, TTS engines, evaluation strategies and UI layers to be independently improved or replaced with disrupting the overall architecture. This modular design increases flexibility for researchers and developers who may wish to extend the system with additional capabilities such as multilingual support, advanced emotion detection, or domain-specific interviewing modes. It also simplifies integration with external services, enabling future versions of the platform to support enterprise-level recruitment tools, learning management systems, or career-support applications. This ensures that the system can evolve adaptability for emerging AI technologies and industry requirements.

V. RESULT AND ANALYSIS

The proposed AI-powered interview system was evaluated across multiple parameters, including candidate engagement, real-time adaptability, and feedback. The platform integrates:

Flask-SocketIO for communication, SQLite for data storage, and AI-driven evaluation models for analyzing candidate responses. Results demonstrate that the system delivers consistent, structured, and rapid assessments that closely mimic real interview environments.

A. Experimental Setup

The platform was deployed locally with the following configurations:

Frameworks: Flask, Flask-SocketIO, SQLite

- AI Models: NLP-based evaluation using word embeddings and cosine similarity for semantic analysis
- Dataset: Simulated candidate responses across both technical and non-technical interview domains

Interviews were conducted for multiple roles, such as Frontend Developer, Backend Developer, HR Screening, and Behavioral Interviews. Each session generated dynamic AI-driven evaluation reports with personalized feedback.

Candidate performance was assessed across three primary dimensions:

- Enthusiasm and Interest
- Communication Skills
- Self-Awareness and Reflection

B. Performance Evaluation

The system highlighted notable improvements over traditional mock interview methods in three key areas:

- Feedback Precision: The AI provided detailed, competency-specific insights, helping candidates identify weak points such as unclear responses or poor self-presentation.
- Real-Time Interaction: Adaptive question flow maintained engagement, replicating authentic interview dynamics.
- Efficiency: Each interview cycle lasted only 5–7 minutes, producing instant structured feedback reports.

A comparison between the proposed and traditional systems is shown below:

TABLE I: COMPARISON OF PROPOSED AND TRADITIONAL INTERVIEW SYSTEMS

Feature	Proposed System	Traditional
Automated Scoring	Yes	No
Real-Time Feedback	Yes	No
Structured Evaluation	Yes	Partial
Time Efficiency	5-7min/report	20-40 min

These results confirm that the system provides a faster, more objective, and scalable evaluation process while maintaining adaptability across both technical and non-technical interview domains.

VI. CONCLUSION AND FUTURE WORK

This system presents a real-time, AI-driven interview agent designed to simulate authentic interview environments while offering structured and adaptive feedback. By leveraging advanced speech and language models, the system delivers consistent and objective evaluation, reduces manual dependency, and enhances learning outcomes for candidates.

Future work includes:

- integrating multilingual support,
- enhancing multimodal emotion recognition,
- optimizing latency for large-scale deployment,
- incorporating gamified learning modules to improve long-term engagement.

The proposed system proved strong potential towards educational institutions, training platforms, and automated hiring solutions requiring scalable and unbiased interview evaluation.

REFERENCES

- [1] J. V. Barpute, S. Pakjade, O. Wattamwar, and S. Diwate, "A Survey of AI-Driven Mock Interviews using GenAI and Machine Learning (InterviewX)," 2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS), 2024.
- [2] Y.-C. Chou, F. R. Wongso, C.-Y. Chao, and H.-Y. Yu, "An AI Mock-interview Platform for Interview Performance Analysis," 2022 10th International Conference on Information and Education Technology (ICIET), 2022.
- [3] T. Sharma, G. Gupta, A. Singh, and S. Singh, "AI-Powered Mock Interview Platform using Computer Vision(CV), Natural Language Processing(NLP) and Generative AI," 2025 3rd International Conference on Self-Sustainable Artificial Intelligence Systems (ICSSAS), 2025.
- [4] C. Clavel, S. d'Armagnac, S. Hebrard, T. Hesters, and D. Potdevin, "Humanized AI in Hiring: An Empirical Study of a Virtual AI Job Interviewer's Social Skills," The International Journal of Human Resource Management, 2025.
- [5] R. M. Marvaniya, V. K. Dabhi, A. S. Acharya, H. B. Prajapati, and D. M. Detroja, "Smart Prep: AI-Based Interactive Interview Preparation System," 2025 4th OPJU International Technology Conference (OTCON), 2025.
- [6] G. ShivajiRao, M. A. Ali, M. Jaiganesh, N. R., M. A. K., and P. K. Parida, "AI-Powered Virtual Job Interview Simulator using NLP," 2025 8th International Conference on Trends in Electronics and Informatics (ICOEI), 2025.
- [7] M. Jagtap, A. Govekar, N. Joshi, S. Joshi, and S. Arora, "AI-Based Interview Agent," Book Chapter, 2022.