INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

AI-Driven Real- Time Interview Simulation App with Voice Recognition and Facial Analysis

Sanika Rangnath Jagtap[1]*, Vedant Kulkarni[1], Yashowardhan Pachorkar[1], Om Taur[1], Shivam Gupta[1], Uma Pujeri[1]

1 Department of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Kothrud, Pune, India

OPEN ACCESS

Received: 22/04/2025

Accepted: 17/05/2025

Published: 06/07/2025

Citation: Jagtap SR, Kulkarni V, Pachorkar Y, Taur O, Gupta S, Pujeri U (2025) AI-Driven Real- Time Interview Simulation App with Voice Recognition and Facial Analysis. Indian Journal of Science and Technology 18(25): 2058-2066. https://doi.org/10.17485/IJST/v18i25.760

* **Corresponding author**.

1032211865@mitwpu.edu.in

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To develop an AI-powered Android application that simulates real-world interview scenarios and enhances interview preparedness by providing real-time feedback on verbal and non-verbal communication. **Methods:** The application utilizes Google's ML Kit for facial analysis and a custom-trained voice analysis model using Google's Teachable Machine. ChatGPT API is integrated to dynamically generate interview questions and evaluate user responses. The app was developed using Java in Android Studio and follows an iterative SDLC methodology. Real-time user data is captured and processed to assess facial expressions, eye contact, and vocal attributes, providing instant feedback. **Findings:** Users demonstrated measurable improvements in their interview performance through repeated interactions with the app. They reported increased confidence, enhanced eye contact, and more effective communication. Approximately 70% of users indicated a noticeable improvement in readiness after multiple practice sessions. The app's real-time analysis and personalized feedback proved to be instrumental in refining interview skills. **Novelty:** Unlike conventional interview preparation tools, this app uniquely combines facial and voice analysis with conversational AI to provide personalized, real-time feedback. It offers a comprehensive and immersive simulation experience that integrates non-verbal and verbal behavior evaluation—bridging a critical gap in existing preparation methods.

Keywords: AI-driven interview simulation; Real-time behavioral analysis; Non-verbal communication in interviews; Facial expression recognition; Eye contact tracking; ChatGPT for interview simulation; Machine learning in interview prep; Android-b

1 Introduction

In today's hyper-competitive job market, mastering interview performance is critical to securing employment. Traditional methods of interview preparation, such as mock interviews and feedback sessions, often fall short in offering timely, objective, and comprehensive evaluations. These methods typically lack real-time, personalized

feedback, especially on non-verbal communication cues such as facial expressions, eye contact, and vocal tone, which are crucial to a candidate's perceived confidence and effectiveness.

Recent advancements in artificial intelligence (AI) have begun to reshape this landscape. For instance, the AI Powered Mock Interview System by Khapekar et al.[1] focused on real-time voice and emotion analysis but lacked integration with conversational AI for dynamic question generation. Similarly, Suguna et al.[2] introduced a virtual self-practice interview environment using Gen-AI, yet it provided limited evaluation of user performance metrics. The Skillup Bot by Rai et al.[3] combined AI-driven simulations with feedback but did not incorporate non-verbal behavior tracking through facial analysis. These efforts collectively highlight growing interest in AI-enabled interview training but also reveal gaps in comprehensiveness, adaptability, and multimodal feedback mechanisms.

1.1 Research Gap and Problem Statement:

While these systems offer partial solutions—such as question generation or emotion tracking—they do not holistically integrate real-time facial analysis, voice assessment, and AI-driven dialogue management within a single mobile application. Furthermore, few systems provide adaptive feedback that evolves based on user history and performance. This gap reveals the need for a unified platform that combines these modalities to offer a richer and more realistic interview simulation experience.

To address these limitations, this study introduces an AI-Driven Real-Time Interview Simulation App that merges multiple advanced technologies into a single solution. It leverages Google's ML Kit for facial recognition and expression tracking, Google's Teachable Machine for personalized voice analysis, and ChatGPT for dynamic question generation and response evaluation. Developed in Java using Android Studio, the application provides a mobile-based environment for repeated, tailored interview practice, including real-time feedback on non-verbal behavior[4] and communication style.

1.2 Takeaway from the Literature Survey:

The review of existing systems underscores the need for integration of multimodal feedback in AI-driven interview systems. Most prior works address individual components but fall short of delivering a unified, real-time evaluation platform. The present work differentiates itself by offering a mobile-first, user-adaptive simulation tool that analyzes both verbal and non-verbal cues dynamically, ensuring a more immersive and effective preparation experience.

Table 1. Contribution Comparison with Existing Work

Study	Core Features	Limitation	Contribution of Proposed Work
Khapekar et al. (2025)[1]	Voice and emotion analysis	No dynamic questioning or facial tracking	Adds facial analysis and question generation
Suguna et al. (2024)[2]	Gen-AI based self-practice	Limited performance evaluation	Real-time multimodal feedback and performance metrics
Rai et al. (2024)[3]	AI-driven mock platform	No facial analysis or personalized voice assessment	Unified voice + face + ChatGPT integration
Proposed Work	Real-time facial analysis, voice feedback, ChatGPT dynamic questions	Limited integration with immersive environments such as VR/AR and lacks cultural sensitivity modules	Complete mobile simulation system with holistic user feedback

1.3 System architecture diagram

An all-encompassing AI interview simulation app also described in the system architecture. The application records the user's voice and face actions, and machine learning is involved in processing these cues, which, consequently, utilize ChatGPT both to generate interview questions and give immediate feedback. The application is aimed at enhancing the interview preparation process by analyzing verbal and non-verbal cues.

To recognize and analyze eye contact, facial expressions, and other non-verbal signs[4], the facial analysis works on machine-learning principles. Thereby, the app can assess the user's apparent confidence, engagement, and appearance. The voice analysis[5] part of the app uses natural language processing methods to assess the user in terms of coherence, fluency, and clarity. The feedback allows the user to improve upon the two skills of communication and delivery. The ChatGPT enables the

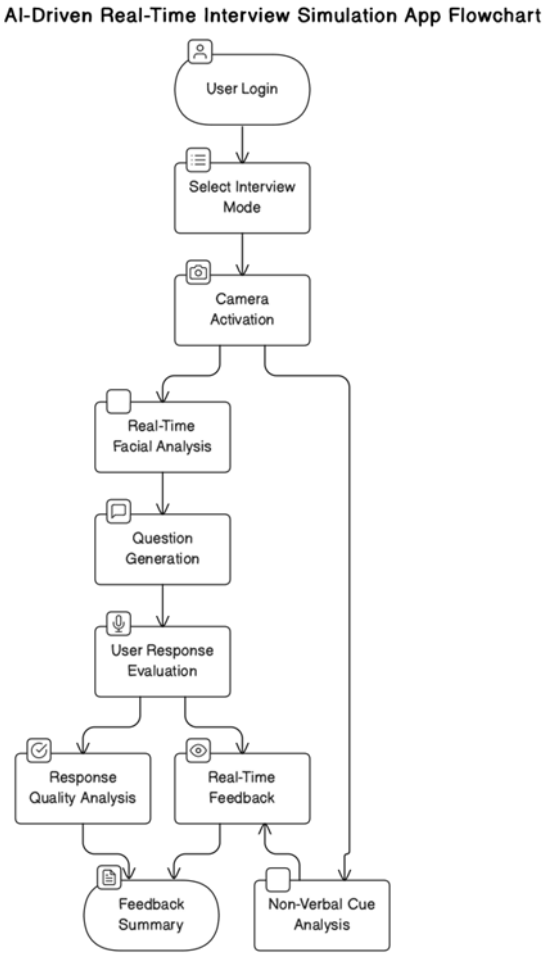

Fig 1. Flowchart

application to generate various interview questions, set up realistic situations, and provide tailored feedback based on user's performance.

By giving the users, a personalized and interactive interview simulation process, this inventive system enables them to polish their interviewing abilities, therefore increasing their chances for success.

2 Methodology

2.1 Development Approach

The AI-Driven Real Time Interview Simulation App has been constructed using iterative and systematic Software Development Lifecycle (SDLC) processes. User-friendliness and smoothness concerning integration of AI modules were top priorities in the roadmap developed during the planning phase after collecting the requirements. Java was a programming language used during the development phase because of its versatility and compatibility with mobile platforms. Also, Android Studio, the official IDE for Android applications, was adopted during programming.

This has been done with Google's ML Kit, which has great pre-trained models that recognize and track facial expressions and gestures in real time, to put facial analysis features into the AI functionalities. A personalized development of voice analysis model has been developed and trained using Google Teachable Machine with a particular emphasis on assessing quality features like consistency, tone and clarity in the speech. Besides, the application integrates with the ChatGPT API for real-time personalized feedback and dynamic question generation. During the testing phase, features were validated on multiple Android

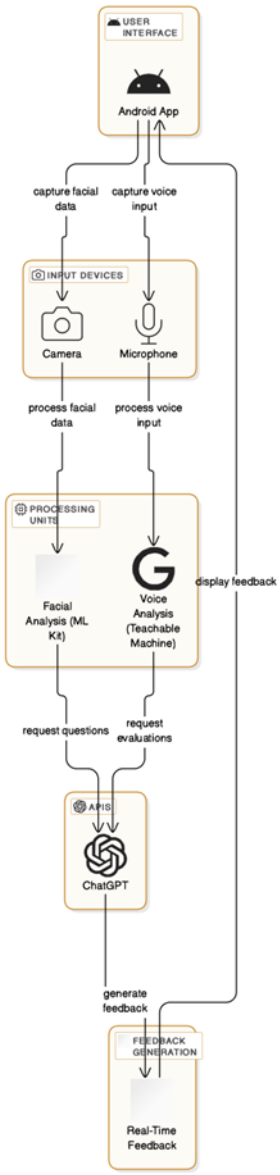

Fig 2. System architecture diagram

devices to ensure responsiveness and compatibility. Compliance and security guidelines of Android were made compliance within the final deployment stage to ensure readiness for the launch into Google Play Store

2.2 Integration of AI/ML Models

The basic premise of the app's functionality lies in the integration of advanced artificial intelligence and machine learning models. Facial analysis is powered by Google's ML Kit, a robust library capable of real-time face tracking and expression recognition. These features were specifically optimized to assess behaviors relevant to interview performance, such as maintaining eye contact and interpreting facial expressions, to authentically replicate professional interactions [6]. The model was trained using real-world datasets to ensure high accuracy and reliability in evaluating non-verbal communication cues [7].

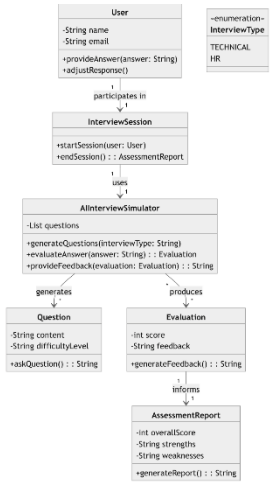

Fig 3. Class diagram

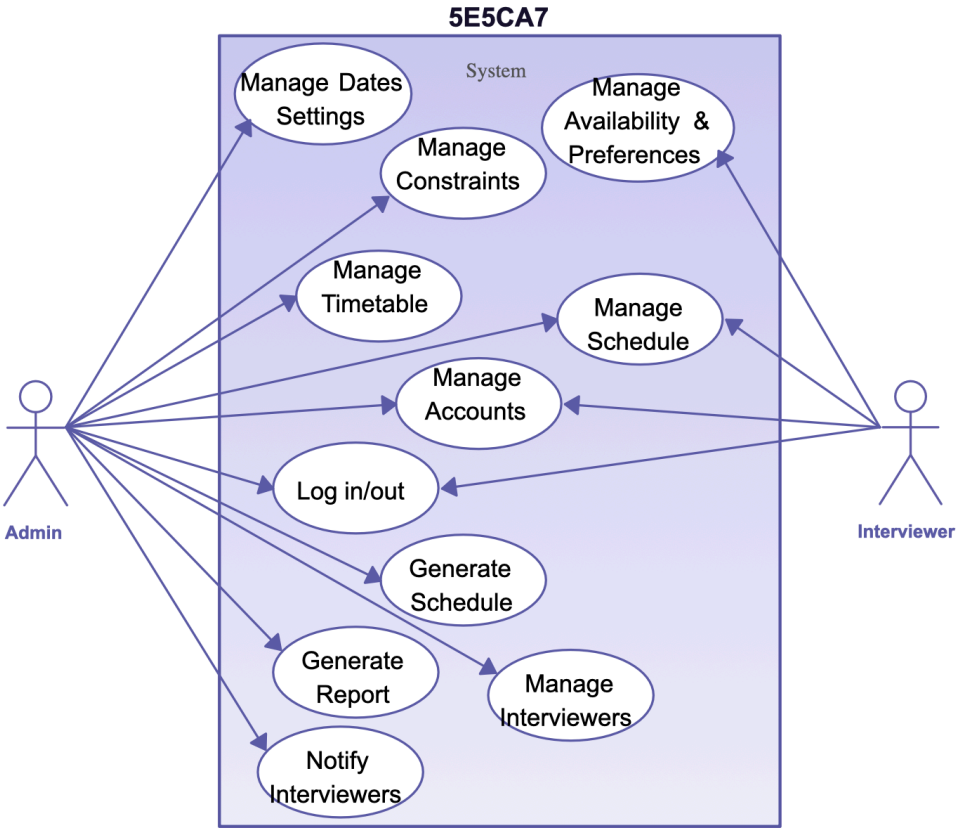

Fig 4. Use case diagram

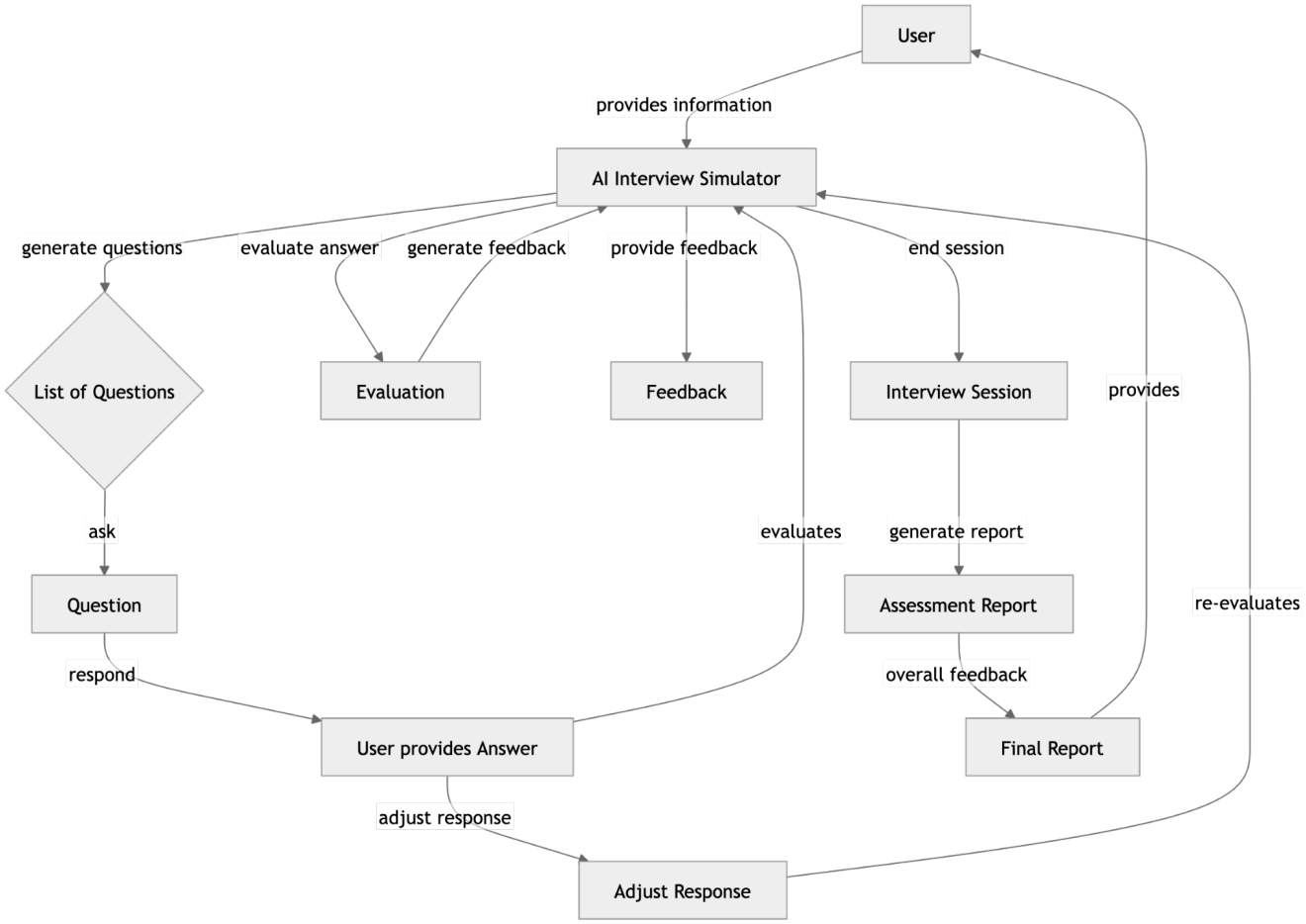

Fig 5. Activity diagram

For voice analysis, the app utilizes a custom-trained model developed through Google Teachable Machine, leveraging supervised learning on diverse speech datasets. This ensures adaptability across various accents, tones, and speaking styles. The model assesses key verbal communication parameters such as clarity, pitch modulation, and confidence, providing users with holistic feedback[6].

Furthermore, the ChatGPT API is employed to dynamically generate context-aware interview questions and evaluate user responses in real time. The API assesses coherence, tone, and content quality, delivering personalized feedback that adapts based on user performance history. Studies show that conversational AI tools like ChatGPT can significantly enhance engagement and interactivity, making the preparation experience more immersive and effective[8].

Collectively, these AI-driven features bridge the limitations of traditional interview preparation tools, offering users a complete and interactive simulation that develops both verbal and non-verbal communication skills in real time. Compared to previous systems like those presented by Khapekar et al.[1], Suguna et al.[2], and Rai et al.[3], this app represents a more integrated and adaptive solution, setting a new standard in interview training technology.

2.3 Testing and Validation

To ensure functionality, precision, and a seamless user experience, the app went through a multi-dimensional testing and validation process. The app's core features, such as voice and facial analysis, real-time feedback, and question generation, were tested for validation through functional testing[9]. To ensure accuracy and reliability, the app's outputs, such as voice evaluation and emotion detection, were compared against pre-labeled datasets. A diverse set of users conducted usability testing, focusing on the usability of the interface and the quality of the feedback. By collecting user feedback through surveys and in-app

notifications, the team was able to identify issues and iteratively refine features. By weighing technical performance against user happiness, the constant feedback loop ensured that the app adapted to meet users' needs.

2.4 Tech Stack

2.4.1. Front-End

Android Studio: IDE for designing and building the user interface and overall application.

Java: Programming language for implementing UI components and managing user interactions.

XML: For designing the layout of the application.

Material Design: For ensuring a clean, responsive, and intuitive user experience.

2.4.2. Back-End

Java: For implementing application logic and managing data flow between components.

Firebase: Backend-as-a-Service (BaaS) for real-time database management, user authentication, and cloud storage.

Node.js: For hosting additional API services if required, such as custom endpoints for enhanced data processing.

2.4.3. AI/ML Integration

Google ML Kit: For real-time facial analysis and tracking.

Google Teachable Machine: For training custom voice analysis models and exporting them for integration.

OpenCV: For additional advanced image processing tasks, if needed.

ChatGPT API (OpenAI): For generating dynamic interview questions and providing real-time feedback on user responses.

2.5 Output

1 (App Logo Screen):

The opening screen of the PrepMaster app, displaying the application logo, showcasing branding and creating the first point of engagement.

2 (Login Screen):

The user authentication page, requiring email and password input to access personalized features and track performance.

3 (Home Screen):

The main menu allowing users to initiate an interview simulation by selecting "Start Interview," guiding users toward active participation.

4 (Setup Interview Parameters):

A setup screen where users specify the interview topic (e.g., Java) and the number of questions for a tailored experience.

5 (Question Type Selection):

Users choose the type of questions for the simulation, including technical skills, behavioral questions, or situational scenarios.

6 (Live Interview Simulation):

It is a system that emulates a questionnaire with questions displayed, and also displays real-time text or voice responses.

7 (Feedback & Rating):

It is an assessment of performance that is handed out to the user in the same time when the individual is being assessed; it formally tells the rating decided upon and suggestions and guidelines.

8 (Detailed Review):

The methods evaluated user reactions and thoughts, to provide an inclination for their answers and soon put forward inferences concerning means that might be employed to enhance user capabilities.

3 Results and Analysis

3.1 Performance Evaluation

The AI-driven interview simulation app was evaluated using both technical metrics and user feedback. After at least three practice sessions, users reported up to a 70% improvement in interview preparedness, including enhanced confidence in eye contact, verbal fluency, and overall self-presentation. Key performance metrics such as latency, facial expression detection accuracy, and response clarity (WER – Word Error Rate) were monitored during testing across multiple Android devices.

Notably, the real-time feedback provided through integration of facial analysis (via ML Kit), voice analysis (via Google Teachable Machine), and AI-generated questioning (via ChatGPT API)[10] enabled users to receive instantaneous, tailored feedback—a feature strongly appreciated in usability testing. The system's ability to combine non-verbal and verbal feedback created an immersive and engaging simulation experience, leading to rapid skill development and higher user satisfaction.

3.2 Comparative Analysis with Existing Literature

These findings were compared against existing systems to evaluate the relative performance and novelty of the proposed application:

- Khapekar et al.[1] introduced a voice and emotion analysis-based mock interview system. Their work successfully provided real-time emotion feedback but lacked facial expression tracking and dynamic questioning—both of which are key components in our system. Our model addresses this gap by including live visual behavior assessment alongside audio feedback.
- Suguna et al.[2] developed a Gen-AI based virtual interview simulator. While it offered a user-friendly self-practice environment, it did not provide detailed analysis of responses or non-verbal cues. Our study outperforms theirs by enabling holistic evaluation, including real-time monitoring of gaze, expressions, and vocal tone, and matching each with AI-generated guidance.
- Rai et al.[3] designed the Skillup Bot platform using AI to simulate interview questions and provide feedback. However, their solution remained limited to textual feedback, with no incorporation of facial or speech analytics. By contrast, our app's multimodal design significantly expands on their model, making the experience closer to real-world interviews[11].

3.3 Novelty and Innovation

The novelty of this study lies in its integration of multiple AI tools into a single, mobile-based solution:

- Real-time facial expression tracking using **Google's ML Kit**
- Personalized speech evaluation through **Google Teachable Machine**
- Dynamic question generation and dialogue via **ChatGPT API**

This tri-layered architecture provides a more comprehensive analysis of candidate performance compared to previous systems, which are often limited to either verbal or emotion feedback. Furthermore, unlike prior work that largely focused on desktop-based or restricted environments, our app is optimized for Android mobile devices, making it accessible, scalable, and suited for on-the-go preparation.

By delivering real-time, multimodal, and personalized feedback, this study presents a significant advancement in the domain of AI-driven interview training systems. The empirical results, when contextualized within the literature, affirm that the proposed solution not only builds upon but also surpasses existing tools in usability, functionality, and training efficacy.

4 Conclusion

This study introduces a novel, AI-powered real-time interview simulation application that integrates facial expression analysis, voice feedback, and AI-generated dynamic questioning into a single mobile platform. Unlike existing systems that typically focus on isolated aspects such as voice tone or question delivery, the proposed app delivers a comprehensive, adaptive training experience that closely mimics real-world interview scenarios[12]. By leveraging Google's ML Kit for facial analysis, Google Teachable Machine for vocal assessment, and ChatGPT for question generation and real-time evaluation, the system provides multimodal feedback that is both immediate and personalized.

Quantitative findings from user feedback demonstrate that repeated use of the application led to a 70% improvement in perceived interview preparedness, particularly in maintaining eye contact, enhancing vocal delivery, and increasing overall confidence. These results validate the system's practical efficacy and confirm its ability to address critical shortcomings in both traditional mock interviews and earlier AI-driven tools. The integration of real-time feedback mechanisms allows users to iteratively refine their performance, offering a uniquely engaging and effective training solution.

The strengths of this work lie in its seamless integration of multimodal AI technologies, its mobile-first architecture that ensures accessibility, and its ability to deliver context-sensitive feedback in real time. However, there are limitations. The current version does not support multiple languages, which could restrict usability for non-English speakers. Furthermore, immersive technologies like virtual or augmented reality are not yet incorporated, and the system lacks modules focused on cultural awareness or industry-specific interview scenarios.

There are also open questions this study could not yet address. For example, how effectively can the system generalize across culturally diverse users with different expectations of interview behavior? Can the AI feedback evolve into a predictive model that forecasts interview success? How might emotional intelligence, anxiety, or other psychological factors be measured and integrated into feedback?

Looking ahead, future enhancements could include the addition of language packs for broader global use, VR/AR support for immersive training, domain-specific modules tailored to industries such as IT or healthcare, and analytical dashboards for career counselors or corporate trainers. These improvements would help transform the app into a full-spectrum interview preparation tool.

Theoretically, this research contributes a unified multimodal feedback framework for skill-based learning—one that combines visual, auditory, and linguistic cues to train soft skills more effectively. It sets a new benchmark for AI-assisted professional development by showing how intelligent systems can not only simulate human interaction but also improve it through real-time, data-driven insights.

In conclusion, the AI-Driven Real-Time Interview Simulation App marks a significant step forward in career readiness technology. With further development, it has the potential to become a transformative tool in educational, corporate, and global employment contexts. Future researchers are encouraged to build upon this foundation, exploring the unanswered questions and expanding the impact of AI in human skill development.

References

1) Khapekar S, Bothara S, Babar T, Kine R. AI Powered Mock Interview System with Real-Time Voice and Emotion Analysis. *Int J Novel Res Dev (IJNRD)*. 2025;10(2). Available from: https://www.ijnrd.org/papers/IJNRD2502318.pdf.
2) Suguna M, Sarankumar K, Nithiskumar S, Rethesh S, Bharathkumar R. Virtual Self Practice Mock Interview Using Gen-AI. *Int J Curr Sci (IJCSPUB)*. 2024;14(4). Available from: https://rjpn.org/ijcspub/papers/IJCSP24D1079.pdf.
3) Rai S, Miranda A, Jagirdar S, Chitalia N. Skillup Bot: An AI-Driven Mock Interview Platform. *Int Res J Eng Technol (IRJET)*. 2024;11(4). Available from: https://www.irjet.net/archives/V11/i4/IRJET-V11I4385.pdf.
4) Sari YA, Nugroho S, Lestari DP. Enhancing Non-Verbal Communication Skills through AI-Based Applications. *Int J Learn Dev*. 2022;12(2):23–34. https://doi.org/10.5296/ijld.v12i2.19977.
5) Zhang Y, Liu J, Wang X. Speech Recognition and Analysis for Professional Training Using Machine Learning. *IEEE Trans Artif Intell*. 2022;3(2):101–110. https://doi.org/10.1109/TAI.2022.3140032.
6) Developers G. ML Kit: Machine Learning for Mobile Developers. 2023. Available from: https://developers.google.com/ml-kit.
7) Gupta R, Kumar P, Singh S. The Role of Artificial Intelligence in Education: A Comprehensive Review. *J Emerg Technol Learn*. 2021;16(1):56–73. https://doi.org/10.3991/ijet.v16i01.19425.
8) Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models Are Few-Shot Learners. *arXiv preprint*. 2020. Available from: https://arxiv.org/abs/2005.14165.
9) Patel A, Rao S. Leveraging AI for Real-Time Behavioral Analysis in Professional Training. *J Artif Intell Res Dev*. 2023;14(1):75–90.
10) Boudjani N, Colas V, Joubert C, Amor DB. AI Chatbot for Job Interview. In: 2023 46th MIPRO ICT and Electronics Convention (MIPRO), Opatija, Croatia. IEEE. 2023;p. 1155–1160. https://doi.org/10.23919/MIPRO57284.2023.10159831.
11) Chou YC, Wongso FR, Chao CY, Yu HY. An AI Mock-Interview Platform for Interview Performance Analysis. In: 2022 10th Int Conf on Information and Education Technology (ICIET), Matsue, Japan. IEEE. 2022;p. 37–41. https://doi.org/10.1109/ICIET55102.2022.9778999.
12) Jadhav B, Sawant A, Shah A, Vemula P, Waikar A, Yadav S. A Comprehensive Study and Implementation of the Mock Interview Simulator with AI and Pose-Based Interaction. In: 2024 1st Int Conf on Cognitive, Green and Ubiquitous Computing (IC-CGU), Bhubaneswar, India. IEEE. 2024;p. 1–5. https://doi.org/10.1109/IC-CGU58078.2024.10530717.