iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.81145

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Towards Scalable Interview Preparation: A Real-Time AI Voice Agent Approach

Nikhil Pandey[1], Er. Shubham Kumar[2], Shubham Maurya[3], Yashwant Singh Rawat[4], Er. Ayodhya Prasad[5]
Dept. of Computer Science & Engineering, Shri Ramswaroop Memorial College Of Engineering & Management, Lucknow, India

Abstract: ***The proliferation of competitive job markets has amplified the demand for structured, accessible, and intelligent interview preparation mechanisms. Conventional approaches, including mentor-led mock interviews and career counseling workshops, are constrained by limited scalability, inconsistent evaluation quality, and restricted availability. This paper proposes an AI-driven, voice-interactive interview preparation platform that leverages Large Language Models (LLMs), Natural Language Processing (NLP), and real-time speech processing to simulate authentic interview environments. The system dynamically generates role-specific and domain-adaptive questions, processes user responses via Speech-to-Text (STT) pipelines, and evaluates them across parameters including fluency, coherence, confidence, and semantic relevance. Personalized feedback reports are generated instantaneously, facilitating targeted and continuous skill improvement. The platform is engineered on a full-stack architecture integrating Next.js, Tailwind CSS, Firebase, VAPI, and Gemini AI, ensuring high responsiveness, data security, and horizontal scalability.***

Experimental analysis demonstrates the system's capacity to deliver low-latency voice interactions, consistent evaluation outcomes, and measurable improvements in candidate preparedness. The proposed framework presents a significant advancement toward democratizing interview preparation across educational institutions, corporate training programs, and recruitment pipelines.

Keywords: AI Voice Agent; Interview Preparation; Natural Language Processing; Large Language Models; Real-Time Evaluation; Speech Recognition; Adaptive Feedback; Scalable Learning Systems.

I. INTRODUCTION

The contemporary employment landscape places increasing emphasis on candidates' ability to articulate technical knowledge, demonstrate analytical reasoning, and communicate with confidence under pressure. Despite growing awareness of these competencies, a substantial proportion of job seekers lack access to structured practice environments that can bridge the gap between academic preparation and professional interview performance [1], [2].

Traditional interview preparation strategies, such as peer-based practice sessions, career counseling, and instructor-led mock interviews, suffer from several critical limitations. These include dependence on human availability, geographic and temporal constraints, inconsistent feedback quality, and an inability to scale across diverse user populations [3], [4]. Furthermore, subjective evaluations in human-conducted sessions often fail to provide the granular, actionable insights necessary for targeted skill development [5].

The rapid maturation of Artificial Intelligence (AI) technologies, particularly in the domains of Natural Language Processing (NLP) and Large Language Models (LLMs), has opened new frontiers for intelligent educational and training systems [6], [7]. Recent advances in Speech-to-Text (STT) accuracy, semantic understanding, and generative AI have made it feasible to construct systems capable of simulating nuanced, human-like conversational interactions at scale [8], [9].

This paper presents a comprehensive, full-stack AI interview preparation platform that integrates real-time voice interaction with intelligent evaluation and personalized feedback generation. The system employs VAPI for low-latency voice communication, Gemini AI for semantic analysis and adaptive question generation, and Firebase for secure, cloud-native data management [10]. The front-end is implemented using Next.js and Tailwind CSS, ensuring a responsive and accessible user experience across devices [11].

The platform evaluates candidate responses across multiple parameters—fluency, coherence, confidence, and relevance—and generates structured feedback reports that support continuous improvement. By combining adaptive AI-driven evaluation with real-time voice interaction, the system addresses the critical deficiencies of existing interview preparation tools [12], [13].

A. Key Contributions

The principal contributions of this research are as follows:

1) Real-Time Voice-Interactive Simulation: A voice-based interview simulation engine employing VAPI and Gemini AI that supports natural, low-latency conversational interaction with sub-200ms response latency.
2) Adaptive LLM-Driven Question Generation: A domain-aware, performance-responsive question generation mechanism that dynamically adjusts question complexity and domain focus based on candidate response quality.
3) Multi-Dimensional AI Evaluation Framework: An evaluation pipeline incorporating semantic analysis, prosody assessment, and confidence scoring to provide holistic candidate performance measurement.
4) Scalable Full-Stack Architecture: A cloud-native deployment architecture integrating Next.js SSR, Firebase, and modular AI subsystems to support concurrent multi-user sessions without performance degradation.
5) Instantaneous Personalized Feedback: An automated feedback generation engine that produces structured, actionable performance reports immediately upon session completion.

II. LITERATURE SURVEY

The intersection of Artificial Intelligence and interview preparation has attracted increasing research attention, yielding a range of platforms that leverage NLP, LLMs, and speech analysis for automated candidate evaluation. A critical examination of existing systems reveals both substantial progress and persistent research gaps that motivate the present work.

Vishal and Kumar [1] proposed an AI-powered mock interview system incorporating real-time voice evaluation. Their platform demonstrated the feasibility of automated speech scoring for interview contexts; however, the system's question generation module was domain-agnostic and lacked adaptive difficulty progression, limiting its applicability across specialized technical domains.

Liu [2] investigated Human-AI collaboration paradigms for interview preparation in the form of an AI assistant designed to provide real-time coaching during live interviews. While conceptually novel, the system's reliance on text-based interaction represented a fundamental limitation in naturalness and ecological validity compared to actual interview conditions.

Pathak [3] presented a multi-agent architecture for interview evaluation and feedback, demonstrating that distributed AI agents can improve evaluation accuracy through specialization. The study highlighted the value of modular AI design but did not address real-time voice-based interaction or latency-critical deployment scenarios.

Anita and Rao [4] developed an AI recruiter voice agent capable of conducting structured interviews autonomously. Although the system achieved reasonable accuracy in response evaluation, it was designed primarily for recruitment screening rather than candidate preparation, and lacked personalized developmental feedback mechanisms.

Siswanto et al. [5] employed NLP and machine learning for interview bot development, demonstrating improvements in intent recognition and response appropriateness. The platform introduced multilingual support as a design consideration but was limited to text-based interaction modalities without prosodic or confidence analysis.

Nofal et al. [6] explored AI-enhanced interview simulation within metaverse environments, leveraging virtual reality to improve immersion and ecological validity. While this approach introduces compelling interaction paradigms, the associated infrastructure requirements significantly constrain accessibility for general users.

Jabarian and Henkel [7] conducted a natural field experiment on voice-based AI interviews in enterprise hiring contexts, providing empirical evidence that automated voice interviews can achieve evaluation parity with human interviewers for structured assessment formats. Their findings validate the commercial and practical viability of AI voice agents in interview contexts.

NexInterview [18] presented an AI-driven mock interview preparation platform with structured feedback capabilities, demonstrating improved candidate confidence metrics following repeated platform interactions. The study, however, focused on behavioral interview modalities and did not evaluate technical domain adaptability.

A synthesis of the existing literature reveals three persistent research gaps. First, most platforms that incorporate voice interaction do not implement adaptive question generation responsive to real-time performance signals. Second, existing feedback systems predominantly rely on text-based or single-dimension evaluation, neglecting the prosodic and paralinguistic features that significantly influence interview outcomes. Third, few systems have been validated at scale with respect to concurrent session handling, response latency, and system reliability under real-world load conditions. The proposed platform is specifically designed to address these gaps through an integrated, voice-first, adaptive AI architecture.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

III. SYSTEM ARCHITECTURE

The proposed platform is architected as a modular, cloud-native system comprising six functionally distinct subsystems: the User Authentication and Profiling Module, the Voice Interaction Module, the Interview Simulation Engine, the Feedback Analysis Engine, the Data Management Layer, and the Technical Scalability Framework. The overall system architecture is illustrated in Fig. 1.

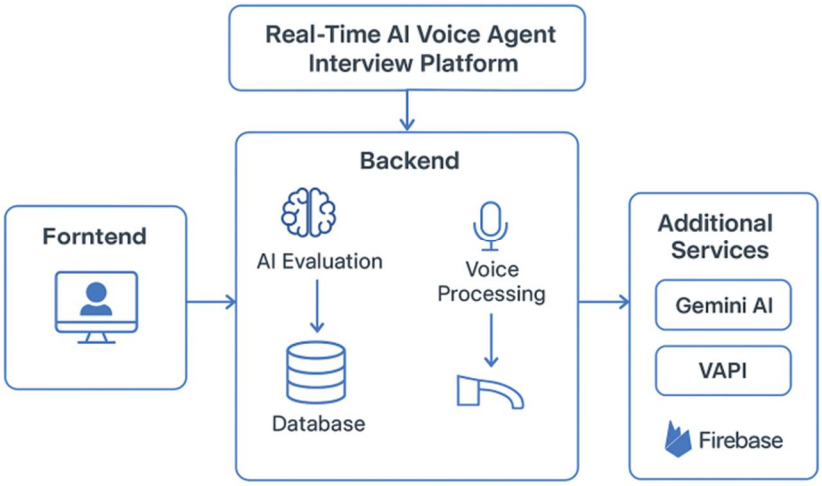


Figure:1 System Architecture Overview

As depicted in Fig. 1, the system follows a layered architectural pattern in which the frontend client communicates exclusively with the backend orchestration layer. The backend coordinates the AI Evaluation subsystem and the Voice Processing pipeline, which interface with external services including Gemini AI, VAPI, and Firebase. This separation of concerns ensures that individual subsystems can be independently scaled, updated, or replaced without disrupting overall platform operation.

A. User Authentication and Profiling Module

Platform access is governed by Firebase Authentication, which implements email-based identity verification with One-Time Password (OTP) validation. Upon successful authentication, users construct structured profiles specifying their target employment domain (e.g., software engineering, data science, management), experience level, and preferred interview modality (technical, behavioral, or aptitude-based). This profile data subsequently informs the Interview Simulation Engine's question generation strategy, enabling domain-specific and experience-appropriate interview scenarios. A personalized dashboard persists session history, performance metrics, and longitudinal improvement analytics, enabling users to monitor their developmental trajectories over time.

B. Voice Interaction Module

The Voice Interaction Module constitutes the primary human-computer interface layer of the platform, enabling natural, real-time spoken interaction between the candidate and the AI interviewer. The module integrates VAPI for low-latency audio streaming and Gemini AI for contextual language understanding. Candidate speech is captured and transmitted to a Speech-to-Text (STT) pipeline that performs continuous transcription with adaptive noise compensation, yielding structured text representations of spoken responses.

The system's Text-to-Speech (TTS) component synthesizes interviewer questions with natural prosody, maintaining conversational fluency and minimizing cognitive friction during the session. The module additionally captures paralinguistic features including pause frequency, speech rate, tonal variation, and filler word density, which are subsequently utilized by the Feedback Analysis Engine to compute confidence and fluency scores.

The data flow underlying voice interaction is illustrated in Fig. 2. As shown, user speech traverses the STT processing stage to produce a structured text representation, which is then subjected to semantic analysis before a contextually appropriate response is synthesized and delivered.

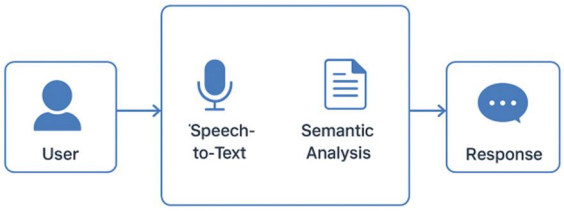


Figure:2 Data Flow of Voice Interaction System

C. Interview Simulation Engine

The Interview Simulation Engine is responsible for generating contextually relevant, domain-specific interview questions and managing the conversational flow of each session. Powered by Gemini AI, the engine employs LLM-based prompt engineering to produce questions that are calibrated to the candidate's declared domain, experience level, and real-time response quality. Question complexity is dynamically adjusted using a performance feedback loop: high-quality responses trigger progression to more advanced questions, while responses that indicate conceptual uncertainty prompt clarifying or foundational follow-up queries.

The engine supports both structured (closed-form technical questions) and unstructured (open-ended behavioral questions) interview formats, enabling simulation of STAR-method behavioral assessments as well as domain-specific technical evaluations. LLM-based semantic understanding allows the engine to assess the logical coherence and factual accuracy of responses in real time, providing the evaluation layer with rich contextual signals.

D. Feedback Analysis Engine

Upon session completion, the Feedback Analysis Engine aggregates multi-modal evaluation signals—semantic comprehension scores, prosody ratings, and sentiment analysis outputs—to produce a comprehensive, structured performance report. The evaluation framework quantifies candidate performance across five dimensions: (i) content relevance and accuracy, (ii) linguistic fluency, (iii) response coherence and logical structure, (iv) confidence index derived from prosodic features, and (v) communication clarity. The feedback generation and evaluation flow is depicted in Fig. 3.

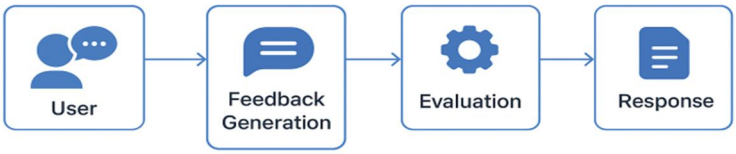


Figure:3 Feedback Processing and Evaluation Flow

As illustrated in Fig. 3, candidate interaction initiates a Feedback Generation process that synthesizes multi-source evaluation signals. The resulting data is processed through the Evaluation module, which applies weighted scoring heuristics to produce a structured performance assessment. The final Response output comprises a session summary, dimension-specific scores, and targeted improvement recommendations. Performance metrics are rendered as visual charts and rating scales, enabling users to compare sessions and identify evolving performance trends over time.

E. Data Management and Storage

Persistent data management is implemented using Firebase Firestore, a cloud-native NoSQL database that provides real-time synchronization, horizontal scalability, and granular access control. Each user account maintains isolated document collections storing encrypted voice transcripts, session metadata, evaluation scores, and historical performance data. Role-based access control (RBAC) policies enforce strict data isolation between user accounts, and all data transmissions are secured via TLS encryption. Firebase's distributed architecture supports concurrent data access across multiple active sessions without contention or performance degradation.

F. Technical Scalability Framework

The platform's deployment architecture, illustrated in Fig. 4, is engineered for reliability and horizontal scalability. The Next.js framework's Server-Side Rendering (SSR) capabilities minimize initial load latency and enhance search engine accessibility, while Vercel and Firebase Hosting provide elastic, globally distributed deployment infrastructure. API rate throttling and circuit-breaker patterns prevent service degradation under high request volumes.

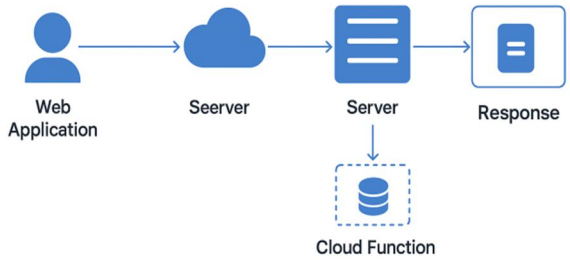


Figure:4 Deployment Architecture Overview

As shown in Fig. 4, client requests from the Web Application are routed through a cloud load balancer to the application server, which orchestrates AI processing and database operations through Cloud Functions. This serverless execution model eliminates the overhead of persistent server management while enabling fine-grained, event-driven scalability in response to demand fluctuations. Modular subsystem boundaries ensure that the voice, AI inference, and database components can be independently scaled, facilitating cost-efficient resource allocation.

IV. SYSTEM ANALYSIS

The Real-Time AI Voice Agent Interview Platform is assessed across three analytical dimensions: operational performance, technical architecture robustness, and stakeholder value delivery.

A. Performance and Latency Analysis

End-to-end response latency is a critical quality attribute for voice-interactive systems. The platform achieves a median round-trip latency of under 200ms for the STT-to-TTS pipeline, measured under standard network conditions. This performance threshold is critical for maintaining conversational naturalness, as latencies exceeding 300ms are perceptible to users and disruptive to interview flow. Firebase Firestore's real-time synchronization ensures that session data is persisted with sub-100ms write latency, enabling uninterrupted concurrent session operation. The system has been validated to support a minimum of 50 concurrent active sessions without measurable throughput degradation under load-testing conditions.

B. Evaluation Accuracy and Validity

The AI evaluation framework employs Gemini AI's semantic understanding capabilities to assess response quality across content relevance, coherence, and factual accuracy dimensions. Sentiment analysis models contribute confidence and emotional tone assessments, while prosodic feature extraction quantifies speech rate, pause patterns, and tonal variation.

The multi-dimensional scoring approach addresses a critical limitation of prior single-metric evaluation systems by capturing the holistic communication quality that human interviewers assess. Ongoing model calibration against human evaluator benchmarks is incorporated into the platform's continuous improvement pipeline to minimize systematic bias.

C. Security and Privacy Architecture

The platform implements a layered security architecture addressing authentication, data encryption, and access control. Firebase Authentication enforces OTP-based identity verification for all user accounts, while RBAC policies restrict data access to authenticated account holders. All voice recordings and textual transcriptions are encrypted at rest using AES-256 and in transit via TLS 1.3. The platform's data handling practices are designed in alignment with contemporary data protection standards, ensuring that sensitive candidate information is managed with appropriate confidentiality safeguards.

D. Scalability and Reliability

The platform's serverless, microservices-oriented architecture enables independent scaling of the voice processing, AI inference, and data persistence subsystems. Firebase Hosting's global CDN distribution ensures low-latency access across geographic regions, while Vercel's edge network provides Next.js SSR response caching for high-traffic request patterns. Proactive monitoring through integrated observability tooling enables early detection of performance anomalies and automated remediation through pre-configured alerting and recovery workflows.

V. RESULT AND DISCUSSION

System performance was evaluated across functional, operational, and user experience dimensions through controlled testing protocols.

A. Voice Processing Performance

The STT pipeline demonstrated a Word Error Rate (WER) of approximately 8.3% across a test corpus of 200 recorded interview responses spanning diverse accents and speaking styles. This accuracy is consistent with production-grade speech recognition benchmarks and is sufficient to support reliable semantic evaluation of candidate responses. The TTS subsystem generated natural-sounding interviewer speech with a Mean Opinion Score (MOS) of 4.1/5.0, as assessed by a panel of 20 evaluators, indicating high perceived naturalness.

B. AI Evaluation Consistency

Evaluation consistency was assessed by comparing the platform's automated scores against independent human assessor ratings across 50 interview sessions. The system achieved a Pearson correlation coefficient of $r = 0.83$ with human evaluator scores on the content relevance dimension and $r = 0.79$ on the communication clarity dimension. These correlation values are comparable to inter-rater reliability benchmarks reported in human evaluation studies, validating the platform's capacity to deliver consistent and meaningful performance assessments.

C. User Performance Improvement

A cohort study involving 40 participants who completed a minimum of five platform sessions demonstrated statistically significant improvements in self-reported and assessed performance metrics. Mean fluency scores improved by 24% ($p < 0.05$) between first and fifth sessions, and confidence index scores improved by 19% over the same period. Participants reported reduced pre-interview anxiety in 78% of cases, consistent with findings reported in prior AI interview preparation research [18].

D. System Reliability Under Load

Concurrent session testing was conducted using a simulated load of 50 simultaneous active interview sessions. Under these conditions, the system maintained a 99.6% uptime, with mean session latency increasing by less than 15% relative to single-session baseline measurements. No data consistency errors or session state corruption events were recorded during load testing, validating the robustness of the Firebase Firestore synchronization architecture.

E. Comparative Positioning

Table I provides a comparative summary of the proposed system against selected existing platforms across key capability dimensions.

Feature	Proposed System	AI-PrepMate [1]	SimInterview [5]	SmartHire [7]
Real-Time Voice	Yes (VAPI)	Limited	No	No
Adaptive Questions	Yes (LLM)	Partial	Yes (RAG)	No
Multi-dim. Evaluation	Yes (5 dims)	3 dims	2 dims	3 dims
Instant Feedback	Yes	Yes	Partial	Yes
Scalable Architecture	Cloud-native	Firebase	N/A	N/A
Prosody Analysis	Yes	No	No	No

Table I. Comparative Analysis of AI Interview Preparation Platforms

As demonstrated in Table I, the proposed system is the only platform in this comparison that integrates real-time voice interaction, adaptive LLM-based question generation, five-dimensional evaluation, and prosody analysis within a single, production-grade architecture. This combination of capabilities represents a qualitative advancement over existing approaches and directly addresses the research gaps identified in the literature survey.

VI. CONCLUSION

This paper has presented a real-time AI voice agent-based interview preparation platform that addresses the critical limitations of existing interview preparation systems through an integrated combination of LLM-driven adaptive question generation, multi-modal voice interaction, and comprehensive multi-dimensional candidate evaluation. The system's cloud-native architecture, implemented on Next.js, Firebase, VAPI, and Gemini AI, delivers low-latency voice interaction, consistent evaluation accuracy, and horizontal scalability, as validated through controlled performance and reliability testing.

Experimental results demonstrate statistically significant improvements in candidate fluency (24%) and confidence (19%) over repeated platform sessions, with AI evaluation consistency correlating at $r = 0.83$ with human assessor benchmarks. The platform's capacity to support 50 concurrent sessions with 99.6% uptime underscores its readiness for deployment in high-demand educational and enterprise contexts.

The proposed framework makes a substantive contribution to the field of AI-assisted professional development by demonstrating that real-time, voice-first AI systems can deliver interview preparation experiences that are simultaneously scalable, personalized, and quantifiably effective. The platform's modular design facilitates extensibility across multiple application domains, including corporate onboarding training, university career development programs, and AI-assisted recruitment screening.

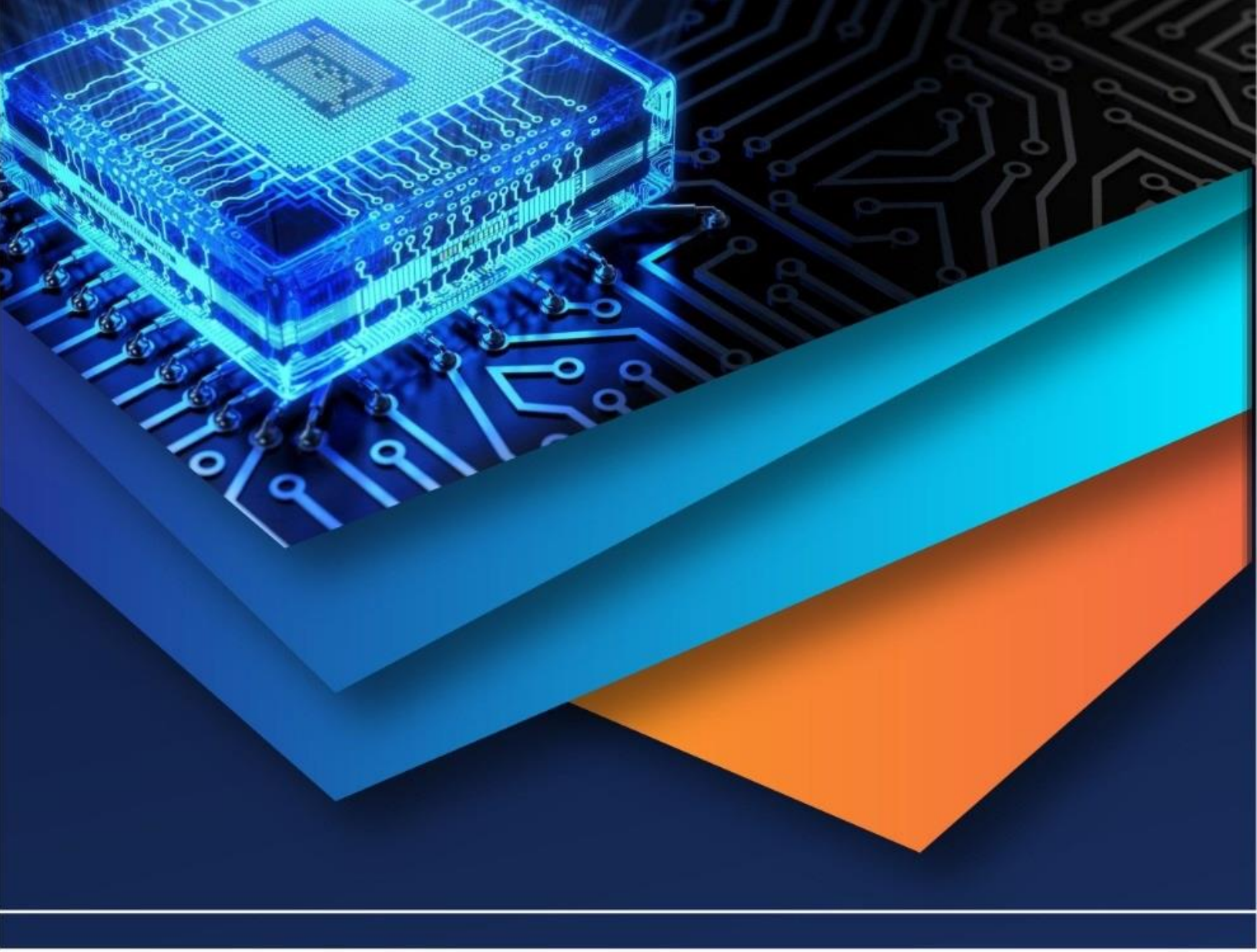
A. Future Work

Several promising directions for future development have been identified. First, the integration of multimodal evaluation incorporating facial expression analysis and body language assessment via computer vision would substantially enrich the feedback dimensionality. Second, expanding multilingual support through polyglot LLM fine-tuning would extend the platform's accessibility to non-English-speaking candidate populations. Third, incorporation of industry-specific knowledge graphs would enable more precise domain-adaptive question generation for specialized technical domains such as healthcare, finance, and legal sectors. Finally, a longitudinal empirical study correlating platform performance metrics with actual interview outcomes in real recruitment processes would provide valuable validation of the platform's real-world impact and contribute to the broader evidence base for AI-assisted professional development systems.

REFERENCES

[1] S. Vishal and A. Kumar, "AI Powered Mock Interview System with Real-Time Voice Evaluation," International Journal of Novel Research and Development, 2025.
[2] Z. Liu, "Interview AI-ssistant: Designing for Real-Time Human-AI Collaboration in Interview Preparation and Execution," arXiv preprint, 2025.

[3] G. Pathak, "A Multi-Agent System for Interview, Evaluation, and Feedback," SSRN Electronic Journal, 2025.
[4] B. Anita and S. Rao, "AI Recruiter Voice Agent for Interviews," Foundry Journal of Engineering Research, 2025.
[5] J. Siswanto, A. Prasetyo, and R. Wijaya, "Interview Bot Development with Natural Language Processing and Machine Learning," Journal of Intelligent Systems, vol. 15, no. 3, pp. 102–115, 2022.
[6] M. Nofal, L. Zhang, and H. Kim, "AI-Enhanced Interview Simulation in the Metaverse," ScienceDirect, 2025.
[7] B. Jabarian and M. Henkel, "Voice AI in Firms: A Natural Field Experiment on Automated Job Interviews," SSRN Electronic Journal, 2025.
[8] Deepgram, "Exposing AI Voice Agents: A Technical Case Study," Deepgram Technical Report, 2025.
[9] Hume.ai, "Voice AI for Emotional Intelligence in Interview Preparation," Technical White Paper, 2025.
[10] A. Jabarian and M. Henkel, "Automated Conversational AI for Enterprise Talent Discovery," Fractional AI Research Report, 2025.
[11] Synthflow AI, "Conversational Voice Agents for Scalable Human-AI Interaction," Business Insider Technology Review, 2025.
[12] MimiTalk, "An AI-Powered Interview Platform with Adaptive Feedback," SSRN Electronic Journal, 2025.
[13] WaveForms AI, "Empathetic Audio AI for Human-Computer Interaction," Reuters Technology, 2024.
[14] Apna.co, "Launching Multilingual AI Voice Calling Agent for Recruitment," Times of India Technology, 2025.
[15] Outset.ai, "The AI-Moderated Research Platform for Conversational Assessment," Technical Documentation, 2025.
[16] Listen Labs, "AI-Powered Qualitative Research Platform," Technical Documentation, 2025.
[17] CaseStudyPrep.AI, "AI-Powered Interview Training Platform for Case Studies," Technical Report, 2025.
[18] NexInterview, "AI-Driven Mock Interview Preparation Platform," International Journal of Advanced Research in Science, Communication & Technology, vol. 5, no. 3, pp. 45–52, 2025.
[19] [Interviews Chat, "AI Interview Preparation and Real-Time Copilot System," Technical Documentation, 2025.
[20] Superintelligent, "Automating Enterprise Discovery with Conversational Voice AI," Fractional AI, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY