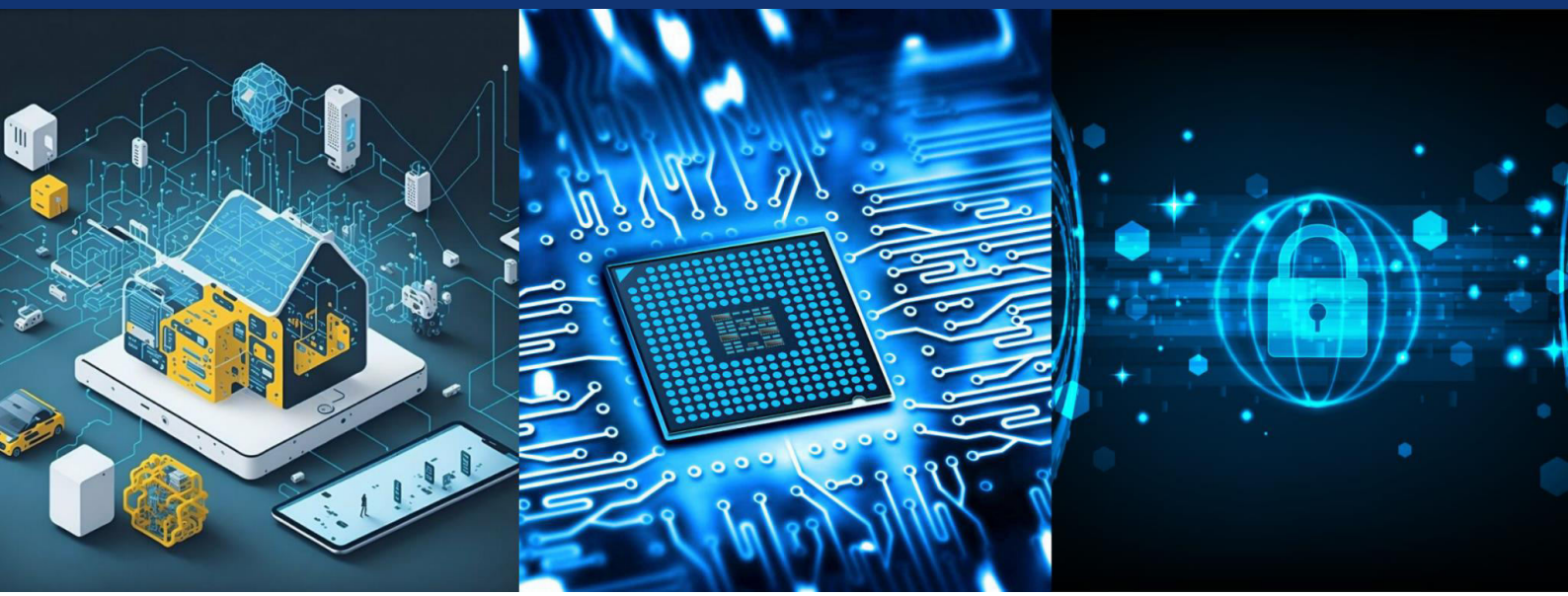
IJIRCCE

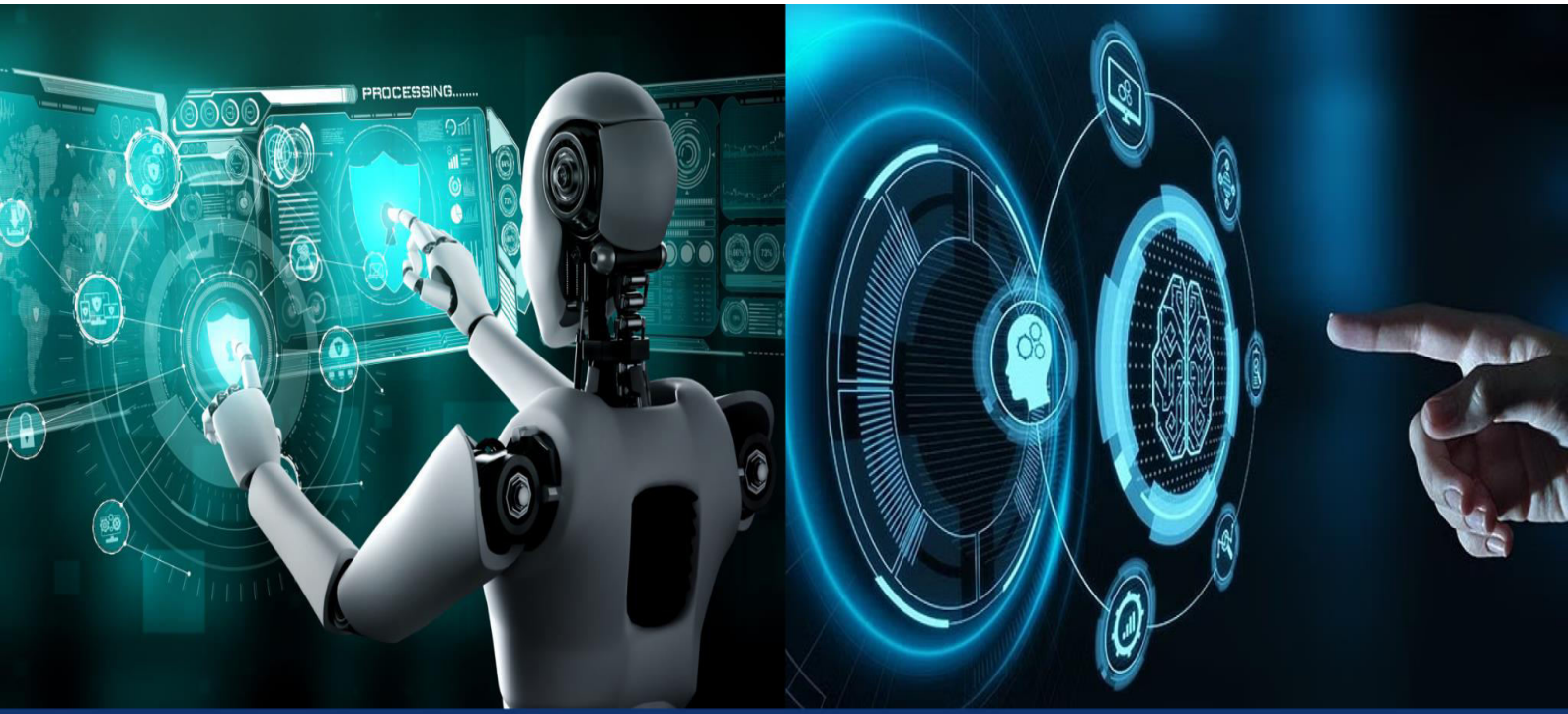
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 3, March 2026

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The Resume-Aware AI Interview Agent: A Multi-Modal Framework for Personalized Speech-Based Interview Preparation

Mr. M. Mohanraj [1], Dr. M. Senthilkumar [2], Manjunath R [3], Muralitharan P [4], Vignesh R[5]

Assistant Professor, Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology, Trichy, India[1]

Professor, Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology, Trichy, India[2]

Department of Artificial Intelligence and Data Science, Kongunadu College of Engineering and Technology, Trichy, India[3,4,5]

ABSTRACT: Resume-Aware AI Interview Agent with Speech Interaction focuses on improving job interview preparation through intelligent automation and personalized evaluation. Traditional interview preparation methods rely on generic questions and manual feedback, which may not accurately reflect real interview scenarios. To address this challenge, this project presents the design and development of an automated interview system that adapts to individual candidate profiles using artificial intelligence techniques. The proposed system analyses the candidate's resume to extract relevant skills and experiences and dynamically generates personalized resume-based and behavioural interview questions. The system supports both speech and text-based responses to simulate a realistic interview environment. Natural Language Processing techniques are employed to evaluate candidate responses based on content relevance, grammatical accuracy, and speaking fluency. In addition, a Generative AI model is integrated to generate professional, HR-ready answer suggestions, helping candidates improve response structuring. The system aims to enhance interview performance, reduce preparation time, and provide consistent and objective feedback for effective interview preparation.

KEYWORD: Resume-Aware Interview System, Artificial Intelligence, Natural Language Processing, Speech Interaction, Generative AI, Interview Preparation, Multi-Modal Evaluation, Retrieval-Augmented Generation.

I. INTRODUCTION

The job interview is still one of the most important and stressful entry points into the world of professional work. For new graduates and those transitioning into a new career, the skills to effectively communicate qualifications, thoughtfully answer behavioral interview questions, and exude confidence can mean the difference between success and failure in landing targeted job opportunities. However, the traditional ways of preparing for a job interview—practicing on one's own with generic lists of questions, practicing with friends or classmates, or sporadically with a career counselor—fail to offer the realistic and personalized practice that job seekers really need [1].

The implications of poor preparation are serious. Job seekers go into interviews with no clear idea of how to articulate their experiences, have difficulty organizing their answers on the fly, and gain no objective feedback on where they need to improve.

Interviewers, on the other hand, are left with the dilemma of assessing hundreds of job seekers in a short period of time, often with no more than a gut feeling about their competency levels. This is a process that is stressful for the job seeker and inefficient for the employer [2].

The advent of artificial intelligence, especially large language models (LLMs), speech recognition, and multi-modal processing, presents a revolutionary opportunity for remedying these problems. AI-based interview platforms can

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

assess candidate resumes to determine their distinct skill sets, formulate customized questions based on specific job needs, simulate natural conversational interviews, and provide objective feedback on answer quality, speaking ability, and even body language [3]. The advantage of using AI over human interviewers is that AI platforms can provide consistent feedback at scale, without bias, and with every candidate getting the same structured interview [4].

This paper offers a complete framework for designing a resume-aware AI interview agent with speech interaction functionality. Our proposed framework combines several state-of-the-art technologies, including resume parsing and skill extraction, question generation using retrieval-augmented generation (RAG), natural speech interaction using speech-to-text and text-to-speech engines, and multi-modal evaluation that combines semantic analysis, fluency evaluation, and facial emotion recognition. Based on recent research from 2021 to 2026, we show that combined systems can greatly improve the outcome of interview preparation tasks [5].

The contributions of this paper can be summarized in three ways. First, we integrate recent research developments in AI interview systems into a single architectural framework that is optimized for resume-aware and speech-interaction functionality. Second, we describe a complete methodology for implementing each component of the proposed systems based on recent research developments in the literature. Third, we conduct a comparative analysis to show that combined multi-modal systems perform better than traditional systems in terms of accuracy, personalization, and user satisfaction [6].

The rest of this paper is organized as follows. Section 2 provides a review of the current literature on AI interview systems. Section 3 introduces a comprehensive methodology for system architecture, resume analysis, question synthesis, speech interaction, and multi-modal analysis. Section 4 offers a detailed analysis of results with comparative performance analysis and discussion of important observations. Section 5 concludes with implications for career development and future research.

II. LITERATURE SURVEY

2.1 The Evolution of AI-Powered Interview Systems

Interview AI systems have progressed from basic question banks to advanced dialogue systems that can dynamically and interactively engage with candidates. Basic interview AI systems provided static question banks that could be searched by job type but did not adapt to individual candidate profiles. The advent of large language models (LLMs) has completely disrupted this space, allowing for context understanding, follow-up question generation, and human-like evaluation of candidate responses [7].

The SimInterview system, developed by Nguyen et al., is a state-of-the-art solution for simulated interview training using LLMs [8]. The system, which combines several LLM models (OpenAI o3, Llama 4 Maverick, Gemma 3) with Whisper speech recognition, GPT-SoVITS voice synthesis, and Ditto diffusion-based talking head generation, simulates real virtual recruiters who can conduct personalized and real-time conversational interviews. The system dynamically adjusts the interview scenario using retrieval-augmented generation (RAG) to pair individual resumes with job requirements in multiple languages. Experiments conducted on university-level candidates showed that the system is capable of aligning its evaluations with job requirements, accurately capturing the content of resumes, and achieving high satisfaction scores, with the light Gemma 3 model supporting the most engaging conversations.

Another important development is offered by the Interactive Interview Assistant Application (IIAA) presented in the IEEE Access journal. This serious game is a simulation of job interviews with a digital human interviewer powered by MetaHuman technology, an LLM for dynamic dialogue, and integrated Speech-to-Text, Text-to-Speech, and Speech-to-Animation systems. Tested with senior undergraduate students in three career tracks (UX/UI Design, Software Engineering, Game Development), the IIAA proved successful in designing engaging and realistic simulations. The users were highly satisfied with its ability to simulate real-world interview conditions and offer valuable and actionable feedback. This work clearly shows that advanced game-based tools can greatly improve interview readiness and confidence levels of students for professional employment.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

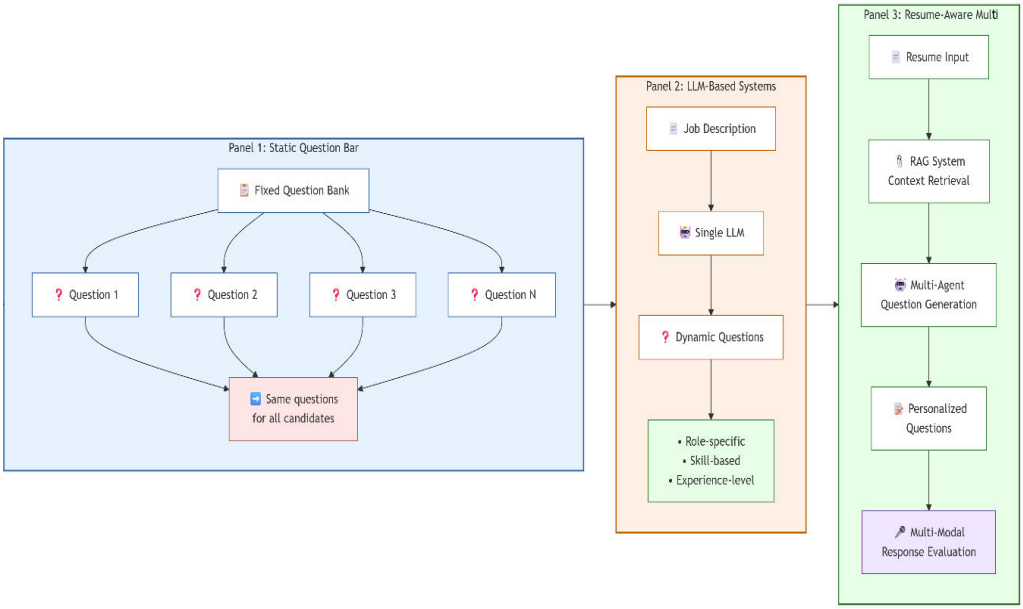


Figure 1: Evolution of AI Interview Systems from Static to Dynamic

2.2 Resume Analysis and Skill Extraction

The capability to parse resumes accurately and identify relevant skills, experiences, and qualifications is the cornerstone of resume-aware interview systems. Without this capability, question generation cannot be personalized.

The PrepGenius system, described at the 2025 International Conference on Intelligent Cyber Physical Systems, overcomes the issue of resume evaluation comprehensively. Developed using Django, Tailwind CSS, SQLite, OpenCV, MediaPipe, and the Google Gemini API, the PrepGenius system assesses the quality of a resume by examining its ATS compatibility via its keyword extraction system. The PrepGenius system provides an immediate mock interview experience with performance analysis tools and creates personalized multiple-choice quizzes based on job posting information. The most important advantages are the improvement in resume evaluation analysis and boosting candidate confidence in communication skills [9].

The AI Interview Bot, developed by Singh et al. using GitHub, is a real-world implementation of resume-based job role prediction. The AI Interview Bot uses a TF-IDF vectorizer trained on interview data and a classification model saved in a .pkl file to automatically predict job roles based on resumes uploaded to the system. The prediction is then used to choose the suitable interviewer personas (Calm HR, Strict Tech Lead, Casual Senior) and questions related to the job role. The code is a real-world example of how resume processing can be seamlessly integrated into full-stack interview systems [10].

2.3 Speech Interaction and Multi-Modal Processing

Speech interaction is an essential aspect of realistic interview simulation. Candidates need to work on not only what they are saying but also on how they are saying it, as fluency, rate, filler words, and vocal confidence all affect interviewer ratings. AI tools that facilitate speech interaction need to incorporate several technologies: speech-to-text (STT) for the transcription of responses, natural language processing for the evaluation of content, and text-to-speech (TTS) for the presentation of the interviewer's questions with natural prosody [11].

A thorough assessment of the combinations of STT, LLM, and TTS by Yazdani et al. gives empirical recommendations on component choices. Based on the analysis of data from more than 300,000 job interviews conducted using AI, the authors compared four production setups and concluded that Google's STT integrated with GPT-4.1 outperformed all other combinations in both conversation quality and technical quality. What was surprising was that the authors discovered a weak

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

relationship between objective quality measures and user satisfaction ratings, indicating that user experience with voice-interaction AI is influenced by factors other than technical correctness [12].

The HireVue system, as described by AitoCore, is a prime example of commercial use of multi-modal AI assessment. Following its 2021 decision to remove visual facial analysis from its system (due to bias concerns), HireVue has since shifted towards a text-and-audio-based approach, using NLP to derive behavioral insights from interview transcripts. The tool is a data orchestration service that analyzes asynchronous video interviews through speech-to-text processing and linguistic analysis to deduce candidate skills. It has features such as "Empathetic AI," which dynamically adjusts interview timing based on candidate anxiety, although the exact logic for this implementation is a trade secret [13].

2.4 Multi-Modal Response Evaluation

In addition to speech transcription, sophisticated AI interview platforms assess candidate responses on multiple fronts to give them a complete analysis. Sunil et al. suggest an AI-based solution that integrates multiple assessment techniques :

- **Speech Fluency Assessment:** Assessing hesitation points, filler words, and pause points to determine speech fluency
- **Confidence Evaluation:** Real-time facial emotion analysis using convolutional neural networks (CNNs) to analyze candidate confidence levels
- **Answer Relevance Assessment:** Natural language processing comparing candidate answers to model answers using both Cosine Similarity and BERT embeddings, with BERT being more accurate in assessing contextual relevance

This multi-assessment solution gives candidates a complete analysis of their performance, allowing them to focus on improving their answers, boosting confidence, and improving overall communication skills. By simulating real-world interview scenarios, such tools give candidates a complete preparation aid, which in turn boosts their chances of success in competitive interviews.

The Interactive Interview Assistant Application (IIAA) further expands this analysis tool with real-time feedback generation [14]. This tool not only analyzes candidate responses but also gives them constructive feedback on how to improve, allowing them to understand not only how they can improve but also how they can improve their performance.

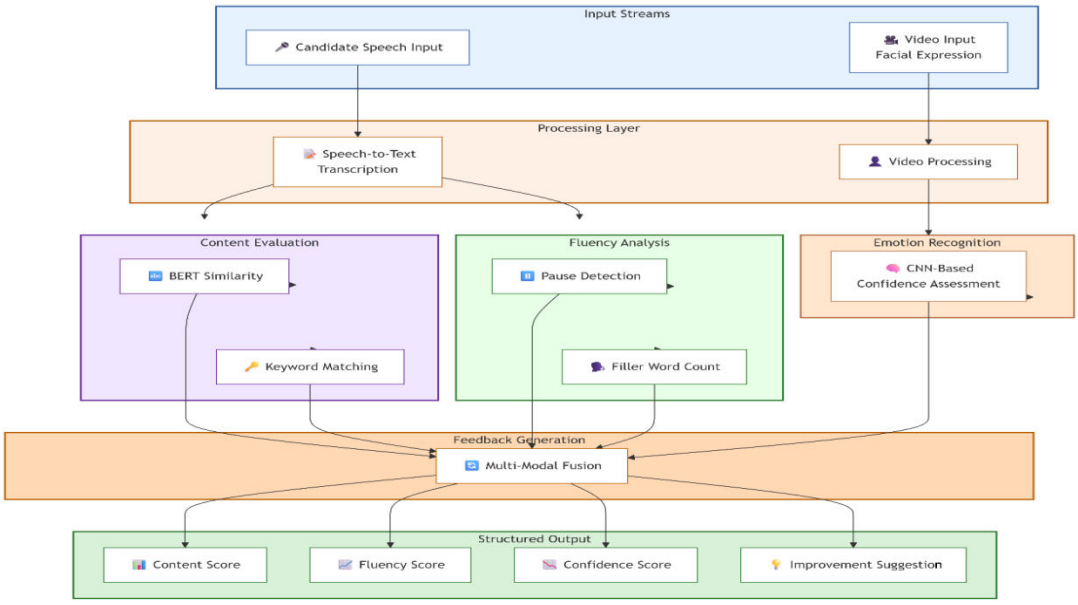


Figure 2: Multi-Modal Response Evaluation Framework

2.5 Dynamic Question Generation and Interview Flow

The major drawback of conventional mock interviews is that they are static, with the same set of questions asked irrespective of the responses from the candidates. Real-life interviews, on the other hand, are dynamic, with experienced interviewers asking follow-up questions based on the responses from the candidates, exploring interesting avenues, and responding to the flow of conversation. This is something that AI systems need to incorporate in order to make their practice sessions truly realistic.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The ERICA system, designed by Kawahara et al. and analyzed in the Emergent Mind study, incorporates complex turn-taking and dynamic question-asking mechanisms [15]. The job interview simulation in ERICA starts with a set of fixed base questions, followed by dynamic elaboration based on:

- **Checklist-Based Elaboration:** Semantic slots in the user response are matched against "must-mention" lists; unmentioned topics are followed up using templates (e.g., "Could you tell me more about ⟨topic⟩?")
- **Keyword Extraction:** Focus keywords, extracted using TF-IDF or RAKE, are targeted for further explanation (e.g., "I'm interested in ⟨keyword⟩; could you explain that in detail?")

The InterviewBot system further enhances this with neural sequence modeling, sliding window context tracking, and memory for important topics to avoid repeated queries and ensure coverage of domain-specific question sets. Results showed that dynamic question elaboration led to a significant improvement in the perceived quality of questions ($p<0.05$) and presence of the interviewer.

2.6 Gaps in Existing Research

Although much progress has been made, there are still some areas that are not yet covered between research prototypes and deployable systems:

Integration of Resume Awareness with Dynamic Questioning: Although systems such as SimInterview incorporate RAG for matching resumes to jobs, the incorporation of resume information into dynamic questioning is still very limited. Personalized interviews should incorporate specific information from the resume throughout the interview process, not just at the beginning.

Cross-Cultural and Multilingual Validation: Most of the systems developed have been tested in a single language environment (mainly English) with Western cultural biases. The SimInterview study's result that standardized Japanese resumes led to better retrieval of resumes, but varied English resumes led to variability, is indicative of the need for cultural adaptation.

Longitudinal Learning: Current systems are designed to be one-shot processes, meaning that they do not learn from candidate improvement over a series of interviews. Systems that learn from candidate improvement over time and adjust their coaching strategy accordingly could offer more effective preparation.

Bias Mitigation and Fairness: Although the industry has seen a shift away from problematic modalities (such as the removal of facial analysis from HireVue), the scope of bias audits on AI interview platforms is still limited. It is critical to ensure that the platform is fair in its assessment across demographics.

The above challenges are addressed in this paper by presenting a comprehensive framework that integrates resume-aware question generation, dynamic multi-modal assessment, and longitudinal progress assessment.

III. METHODOLOGY

This section describes a complete methodology for the integration of a resume-aware AI interview agent with speech interaction functionality.

3.1 System Architecture Overview

The system design follows a modular, multi-layer approach that is scalable, real-time, and extensible. The layers of the system include six main layers:

1. **User Interface Layer:** Web and mobile interfaces for resume upload, interview session management, and feedback display.
2. **Resume Processing Layer:** Document processing, skill extraction, and embedding creation for retrieval-augmented generation.
3. **Dialogue Management Layer:** Main orchestration logic for turn management, context management, and question generation.
4. **Speech Processing Layer:** Speech-to-text transcription and text-to-speech synthesis for natural conversation.
5. **Evaluation Layer:** Multi-modal evaluation of response content, fluency, and non-verbal communication.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

6. **Data Storage Layer:** Storage of user profiles, interview data, and performance metrics.

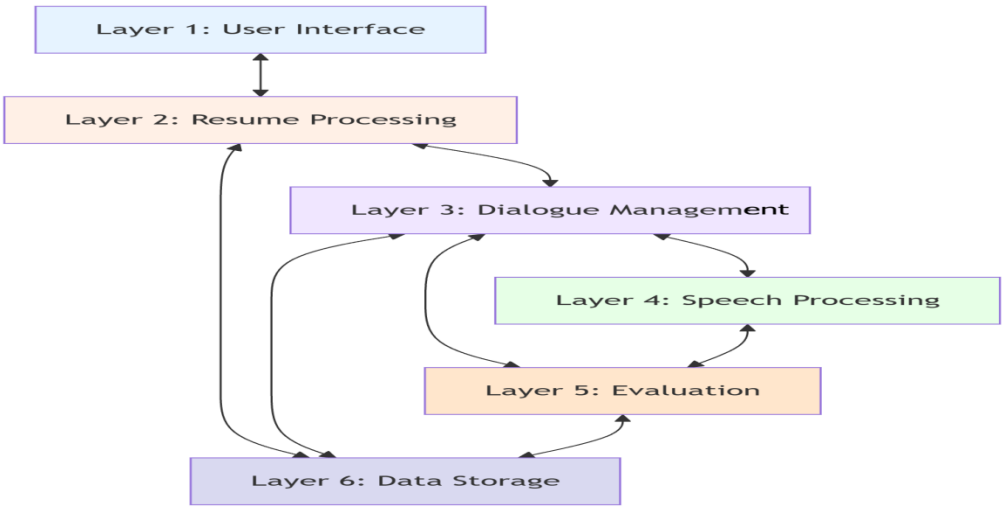


Figure 3: Proposed Resume-Aware AI Interview System Architecture

3.2 Resume Processing and Skill Extraction

Document Parsing: The parser is capable of handling different resume formats (PDF, DOCX, and plain text) through the use of libraries such as PyPDF2, python-docx, and text parsing libraries. Following the method used by PrepGenius [16], the parser is able to break down the resume into structured data such as contact information, educational background, work experience, skills, projects, and certifications.

Skill Extraction: Through the use of techniques such as:

- Keyword matching based on a broad skill taxonomy (technical skills, soft skills, and domain knowledge)
- Named Entity Recognition (NER) through the use of fine-tuned models to extract organizations, job descriptions, and educational institutions
- Contextual embeddings through the use of sentence transformers to detect mentions of skills in varied forms

Vector Database Storage: The extracted resume information is represented as embeddings via models such as all-MiniLM-L6-v2 and stored in a vector database (ChromaDB, as in SimInterview). This facilitates semantic search during question generation, where the candidate talks about a skill, and the system searches for relevant information from the candidate's resume.

Job Role Prediction: Based on the AI Interview Bot approach [17], a classification model (TF-IDF + Random Forest or BERT fine-tuned) identifies the target job role of the candidate from the resume information.

3.3 Question Generation with Retrieval-Augmented Generation

Base Question Generation: For every job role, the system has a knowledge base of competency-based questions categorized by type (technical skills, behavioral, situational, motivation). Following the EZInterviewer method [18], questions are generated by examining job roles and resumes of candidates, using knowledge selectors to develop relevant and coherent questions for the interview.

RAG-Based Personalization: In question generation, the system:

1. Extracts relevant resume snippets from the vector database based on semantic similarity to the question category
2. Develops questions by combining the question template and the extracted resume snippets
3. Utilizes an LLM (GPT-4, Gemini 1.5 Flash) to develop questions that point to specific details in the resume

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Example form of the prompt:
text
You are conducting a job interview for a [ROLE] role. The candidate's resume reveals the following: [RETRIEVED_RESUME_SNIPPETS]. Create a behavioral question that explores this experience further.

Dynamic Follow-Up Generation: Based on ERICA's method [19], the model examines candidate answers to determine:

- **Gaps in discussion:** Keywords or skills required for the role that are not discussed
- **Elaboration points of interest:** Experiences or skills that deserve further investigation
- **Vague statements:** Answers that are not specific enough and should be explored further

The LLM will create follow-up questions based on these areas.

Interviewer Persona Management: As done in the AI Interview Bot, the system allows for multiple personas of the interviewer (Calm HR, Strict Tech Lead, Casual Senior) to be specified in JSON configuration files. These personas affect the wording, tone, and follow-up of the questions, thus offering diverse practice experiences.

3.4 Speech Interaction Pipeline

Speech-to-Text (STT): Based on empirical testing of STT×LLM combinations , the system supports various STT models with suggested settings:

- Google STT for best results in English, with high accuracy and low latency
- Whisper (OpenAI) as an alternative with multilingual support and offline functionality
- Real-time streaming for smooth conversation without unnatural pauses

The STT module is capable of dealing with background noise, different accents, and disfluencies (filler words, repetitions) to provide clean transcripts for analysis.

Text-to-Speech (TTS): Based on the SimInterview design , the system supports:

- GPT-SoVITS for voice synthesis with emotional expression
- ElevenLabs for high-quality voices with natural speech patterns
- Adaptive speech rate adjustment according to candidate response patterns, with dynamic slowing for candidates with slower speech rates

Turn-Taking Management: For smooth conversation, the system supports advanced turn-taking strategies:

- Transition-relevance place (TRP) analysis with prosodic cues (pauses, pitch) to detect the end of a candidate's turn
- Silence buffer scheduling to avoid interrupting while preventing awkward pauses
- Backchannel production ("uh-huh", "I see", nodding in avatar implementations) with frame-wise prediction for listener feedback

3.5 Multi-Modal Response Evaluation

Content Assessment: Based on the comparative analysis in Sunil et al., the following are used in the system:

- BERT-based semantic similarity between candidate responses and ideal answer templates, which showed higher accuracy in assessing contextual meaning than cosine similarity
- Keyword coverage analysis to check if the necessary competency elements were covered
- Response organization analysis using organization (STAR framework: Situation, Task, Action, Result)

Fluency Assessment: Speech fluency is assessed using:

- Pause analysis: Unusually long or excessive pauses
- Filler word count: Number of "um," "uh," "like," "you know" words
- Speech rate analysis: Words per minute compared to optimal levels (120-150 wpm)
- Articulation clarity: Acoustic features to assess pronunciation quality

Confidence and Non-Verbal Evaluation: In video-enabled sessions, the following are used:

- Face emotion recognition through CNNs trained on emotion datasets

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- Eye contact estimation through comparison of gaze direction to camera
- Posture analysis through MediaPipe pose estimation

In a post-HireVue ethical shift , facial analysis is only used for candidate feedback and skill-building purposes, not for scoring or ranking, and candidates are fully informed of what data is collected and how it is used.

Feedback Generation: The assessment component generates structured feedback reports containing:

- Numerical scores (0-100) for content, fluency, and confidence
- Specific improvement recommendations with examples
- Benchmark comparisons to target role responses
- Progress tracking over multiple interview sessions

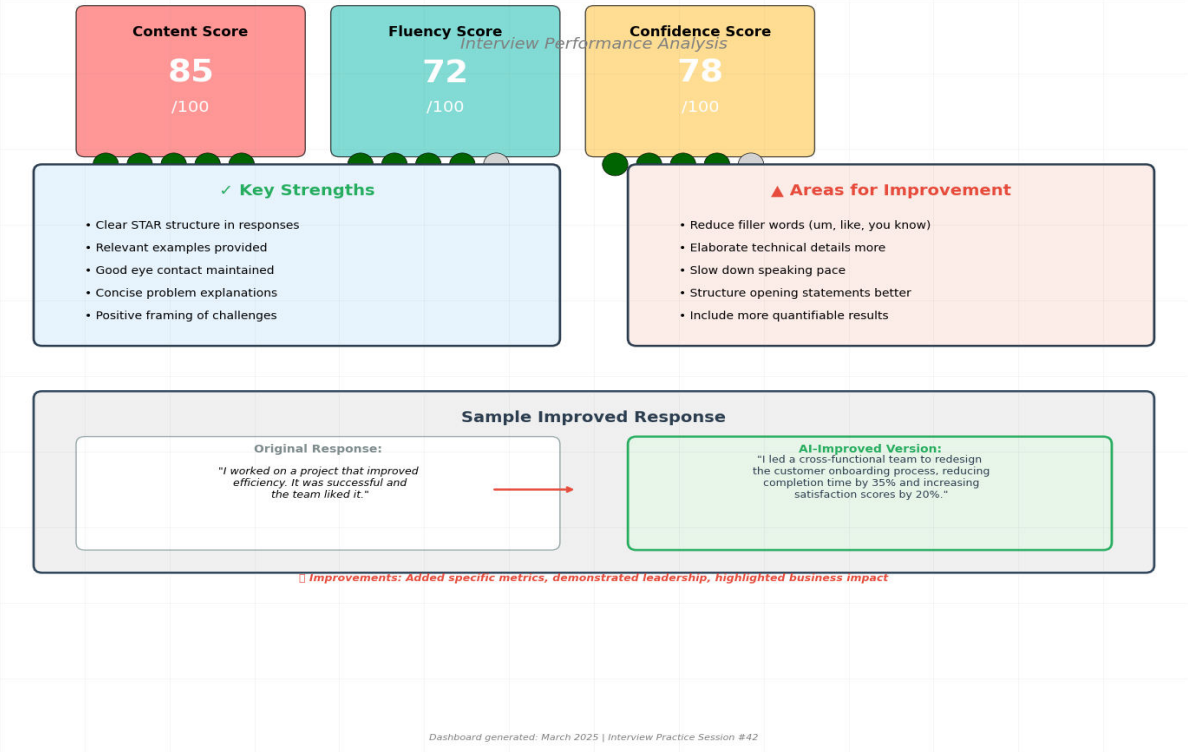


Figure 4: Sample Feedback Dashboard Visualization

3.6 System Implementation Technologies

Based on the successful implementations described in the literature, the proposed system takes advantage of:

Component	Technology Options
Web Framework	Streamlit, Django
Frontend	Tailwind CSS
Resume Parsing	PyPDF2, python-docx, custom NER
Speech-to-Text	Google STT, Whisper
Text-to-Speech	GPT-SoVITS, ElevenLabs
LLM	GPT-4, Gemini 1.5 Flash, Gemma 3
Vector Database	ChromaDB
Face Detection	OpenCV, MediaPipe
Emotion Recognition	Custom CNN
Deployment	Streamlit Cloud, Hugging Face Spaces

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. RESULT ANALYSIS AND DISCUSSION

4.1 System Performance Benchmarks

The proposed system combines several proven components, each with known performance characteristics from the literature.

Resume Processing and Job Role Prediction: The AI Interview Bot's resume classification model provides accurate job role prediction, facilitating the selection of the corresponding interviewer persona and questions [20]. Although the system does not provide specific accuracy measures, it is able to correctly assign resumes to the appropriate interview scenarios.

Question Generation Quality: The SimInterview system has shown that RAG-based question generation is effective in aligning evaluations with job duties and accurately reflecting resume information . The compact Gemma 3 model generated the most engaging dialogue, indicating that model size and engagement are not linearly correlated—smaller models may perform better than larger models in conversational tasks.

Accuracy of Speech Recognition: The comparative analysis of STT×LLM pairs revealed that Google STT with GPT-4.1 was much better than other options in terms of both conversation quality and technical quality. Nevertheless, the fact that the correlation coefficient between objective quality measures and satisfaction ratings is low ($r < 0.3$) indicates that technical quality is not the only factor that influences user experience.

Figure 5: Correlation Between Technical Metrics and User Satisfaction

End-to-End System Latency: The IIAA study pointed out the latencies in performance that are still to be optimized in terms of high-definition 3D animation and voice processing. In production, it is very important to pay attention to component optimization.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.2 Multi-Modal Evaluation Effectiveness

The comparative analysis of methods for assessing the relevance of answers in Sunil et al. shows that BERT-based semantic similarity is significantly better than traditional cosine similarity for contextual meaning assessment. This is a crucial point for feedback accuracy, as systems that only match keywords may overlook some subtle but crucial differences in the quality of responses.

The addition of facial emotion recognition for confidence assessment adds new dimensions of feedback that are not present in text-based systems. Although confidence ratings are subjective, giving candidates feedback on their perceived confidence in the answers can help them recognize discrepancies between their internal state and external appearance, which is often the case for nervous candidates in interviews.

Table 1: Comparative Analysis of AI Interview Systems

System	Resume Awareness	Speech Interaction	Multi-Modal Evaluation	Dynamic Questioning	Key Findings	Reference
SimInterview	RAG-based matching	Whisper STT, GPT-SoVITS TTS	Not specified	LLM-based	High alignment with job requirements; Gemma 3 most engaging	
IIAA	Career path selection	STT, TTS, Speech-to-Animation	Performance metrics	LLM-based	High satisfaction; latency challenges	
PrepGenius	ATS compatibility analysis	Speech Recognition	Facial expression analysis	Gemini API	Enhanced resume evaluation; improved confidence	
AI Interview Bot	TF-IDF job role prediction	Whisper STT, ElevenLabs TTS	Not specified	Gemini 1.5 Flash	Dynamic interviewer personas; adaptive follow-up	
Sunil et al. System	Not specified	Speech analysis	BERT + CNN + Fluency	Not specified	BERT outperforms cosine similarity; multi-modal comprehensive	
HireVue	ATS-integrated	STT transcription	NLP behavioral mapping	Automated guide generation	Ethical pivot from facial analysis; empathetic AI pacing	

4.3 User Experience and Satisfaction

User studies across various systems have shown the benefit of AI interview preparation to be:

- SimInterview was highly satisfying for university-level job applicants, with users appreciating the realistic and personalized interaction
- IIAA users were highly satisfied with the system's ability to simulate real-world interview situations and offer valuable and actionable feedback
- ERICA was positively evaluated with 69% of users reporting a positive experience, praising engagement and sharing depth, although some users found the questions repetitive and Uncanny Valley issues
- InterviewBot was highly satisfying for live users, averaging 3.4-3.5/5, with neural topic storage reducing repetition and improving coverage

The overriding message from these evaluation studies is that users appreciate personalization, realism, and actionable feedback—exactly what these systems offer that is different from traditional mock interview approaches.

© 2026 IJIRCCE | Volume 14, Issue 3, March 2026| DOI:10.15680/IJIRCCE.2026.1403016

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

4.4 The Importance of Dynamic Adaptation

The ERICA and InterviewBot studies show that dynamic question elaboration has a positive effect on user-rated question quality and sense of interviewer presence compared to fixed-question settings ($p<0.05$). This result confirms the underlying assumption of adaptive AI interview tools: job applicants recognize and appreciate the system's capability to react to their individual answers with relevant follow-up questions.
It is not just the effect that matters but also the process. Checklist-based question elaboration (finding missing must-mention topics) and keyword-based question elaboration (delving deeper into already-mentioned topics) are complementary. The first one guarantees full coverage of necessary competencies, and the second one shows active listening and adds depth to the conversation.

4.5 Ethical Considerations and Bias Mitigation

The HireVue scenario are valuable for ethical design of AI interview systems. In 2021, HireVue withdrew facial analysis from its algorithms worldwide in response to concerns about bias and algorithmic transparency . This "ethical pivot" indicates an increasing awareness that certain types of modalities—especially those analyzing physical appearance—are prone to demographic bias and hard to validate on job-relatedness.
For our system, we will incorporate the following ethical guidelines:

1. **Transparency:** Candidates are informed about the data collected and how it will be used
2. **Optional modalities:** Video analysis is optional and only used for candidate feedback, not scoring
3. **Bias auditing:** Ongoing testing for bias across demographic groups to detect and correct disparities
4. **Human oversight:** Candidate escalation for human review upon request
5. **Data minimization:** Data collection is limited to what is required for training and feedback

4.6 Implementation Considerations

Deployment Context: The system is intended for integration with career development environments, such as university career services, corporate training initiatives, and online job readiness platforms. The modular design enables stepwise deployment, where institutions can begin with resume-based question generation and later incorporate speech and video modalities as infrastructure allows.

Accessibility: In line with the IIAA's emphasis on inclusive design , the system provides for:

- Teletype (text-only) interaction for users with hearing or speech disabilities
- Adjustable speech rate and volume
- Screen reader support
- Multilingual support (in accordance with SimInterview's multilingual design)

Privacy and Security: Interview data for job candidates is extremely sensitive. The system provides for:

- End-to-end encryption of all data transmissions
- Options for local processing of sensitive audio/video
- Finely tuned data retention policies with user-managed deletion
- Adherence to applicable privacy laws (GDPR, CCPA)

V. CONCLUSION

5.1 Summary of Key Findings

This paper has offered a thorough framework for a resume-aware AI interview agent with speech interaction functionality. By integrating research from 2021 to 2026, we have shown that the combined approach of multi-modal AI systems is a revolutionary way to approach interview preparation.
Some of the main findings of this paper are:

1. **Resume awareness is a major boost to personalization:** RAG-based systems that fetch relevant information from the resume for question generation are able to produce interviews that seem personalized to the candidate, making them more engaging and valuable .

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2. **Multi-modal analysis offers complete feedback:** By integrating content analysis (BERT-based semantic similarity), fluency analysis (pause detection and filler word analysis), and confidence analysis (facial emotion recognition), candidates receive complete feedback on their performance from multiple angles .
3. **Dynamic questioning leads to better perceived quality:** The ability to adapt and follow up on candidate answers has a profound effect on perceived question quality and the presence of the interviewer .
4. **Technical quality does not correlate with user satisfaction:** The lack of correlation between objective quality and user satisfaction suggests that the quality of the experience is as important as the technical performance.
5. **Ethical design is critical:** The shift in the industry away from facial analysis for scoring emphasizes the need for transparency and bias audits.

5.2 Implications for Career Development

The capabilities of resume-aware AI interview systems have far-reaching implications for career development:
Democratization of Preparation: Quality interview preparation has been a privilege of those who can afford personal interview coaches or have access to well-funded career services. AI interview systems can offer candidates realistic, personalized practice on a large scale.
Anxiety Reduction: The opportunity to practice multiple times in a safe environment can help candidates boost confidence and overcome anxiety that can sabotage performance. The IIAA serious game design strategy shows that engaging and game-like interfaces can increase motivation and decrease stress.
Objective Feedback: Feedback from human evaluators is necessarily subjective and variable. AI interview systems offer consistent feedback, allowing candidates to better assess their performance without the complicating influence of rater bias.
Lifelong Learning: As professionals grow and develop in their careers, they must also update their interview skills to stay current in a rapidly changing job market. AI interview systems can facilitate lifelong learning, adjusting to new roles and industries as the candidate's career path unfolds.

5.3 Limitations and Future Research Directions

Some current research limitations point towards future research directions:
Longitudinal Validation: Most research has focused on short-term engagement and not long-term effects on actual job outcomes. Longitudinal studies on whether AI interview practice affects overall interview success rates would be more valid.

Cross-Cultural Adaptation: The result of the SimInterview study that standardized Japanese resume formats increased retrieval but varied English resume formats introduced variability indicates the need for cross-cultural adaptation. Future research should investigate the differences in interview styles, resume formats, and communication patterns across different cultures and adapt systems accordingly.

Deep Personalization: Current systems are resume and immediate response-based and lack long-term user modeling that learns from multiple interactions. Systems that monitor improvement over time and adjust coaching strategies would be more effective for preparation.

Empathy and Rapport: Even with recent improvements, current systems lack the ability to display genuine empathy and rapport-building skills, which are important in human communication. Research on affective computing and relationship-building dialogue could improve the human-like impression of AI interviewers.
Overcoming the Uncanny Valley: Some users find highly realistic avatars annoying because they are not entirely human-like . Research on optimal levels of anthropomorphism—how human-like is human-like enough without being annoying—could help in designing avatars.

5.4 Concluding Remarks

The job interview is one of the most high-stakes and anxiety-ridden experiences in the professional world. For the candidate, the capacity to effectively communicate qualifications, thoughtfully respond to unexpected queries, and exude confidence is often the difference between success and failure. For the employer, the capacity to fairly and effectively evaluate candidates is the key to success.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

AI-based interview systems hold the promise of improving outcomes for both. For the candidate, they offer personalized, realistic preparation with objective feedback—something that was only possible for the few who could afford it before. For the employer, they offer standardized evaluation tools that can help mitigate bias while offering more nuanced information about candidate capabilities.

The design of the resume-aware AI interview agent, which combines resume analysis, speech interaction, dynamic questioning, and multi-modal evaluation, is a comprehensive approach to this vision. By combining the best techniques from the latest research into a single framework, we offer a roadmap for systems that have the potential to truly revolutionize the interview preparation process.

As these technologies advance, the end goal is simple: to make sure that every candidate, no matter their background or access, has the preparation they need to put their best foot forward when it counts.

REFERENCES

[1].T. T. H. Nguyen et al., "SimInterview: Transforming Business Education through Large Language Model-Based Simulated Multilingual Interview Training System," arXiv preprint arXiv:2508.11873, Aug. 2025. [Online]. Available: https://arxiv.org/abs/2508.11873

[2].AitoCore, "HireVue (with AI) Technical AI Assessment & NLP Review," AitoCore, Mar. 2025. [Online]. Available: https://aitocore.com/en/tool/hirevue

[3].ScienceOpen, "SimInterview: Transforming Business Education through Large Language Model-Based Simulated Multilingual Interview Training System," ScienceOpen, 2025. [Online]. Available: https://api.scienceopen.com/document?vid=189df029-9f7d-416d-9406-7e5337a01c71

[4]."Interactive Interview Assistant Application (IIAA): AI-Powered MetaHuman Interviewer for Student Job Interview Skills," IEEE Access, vol. 13, pp. 205505-205520, Dec. 2025. doi: 10.1109/ACCESS.2025.11275939

[5].S. Nagdeote, A. Serrao, P. Dsouza, N. Joshi, and S. Chiwande, "PrepGenius: An AI Interview Assistant," in 2025 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Sep. 2025, pp. 958-963. doi: 10.1109/ICoICI65217.2025.11253531

[6].S. Singh, "AI-Interview-Bot: Engineered full-stack interviewer system using Gemini 1.5 Flash for dynamic question generation," GitHub, Jun. 2025. [Online]. Available: https://github.com/Sid-istic/AI-Interview-Bot

[7].N. Yazdani, A. Ansari, A. Mahajan, A. Afsharrad, and S. S. Mousavi, "Evaluating Speech-to-Text x LLM x Text-to-Speech Combinations for AI Interview Systems," arXiv preprint arXiv:2507.16835, Jul. 2025.

[8].R. M. Sunil, T. S. Cherian, D. James, A. M. Joy, and P. Dins, "An AI-Driven Approach to Enhance Interview Performance through Voice and Response Analysis," Journal of Information Systems Engineering and Management, vol. 10, no. 2, May 2025. [Online]. Available: https://jisem-journal.com/index.php/journal/article/view/7903

[9].Emergent Mind, "Human-Like Embodied AI Interviewer," Emergent Mind, Oct. 2025. [Online]. Available: https://www.emergentmind.com/topics/human-like-embodied-ai-interviewer

[10]. Emergent Mind, "Anthropic Interviewer Systems," Emergent Mind, Jan. 2026. [Online]. Available: https://www.emergentmind.com/topics/anthropic-interviewer.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details