Parikshit Jaybhaye, 2026, 14:1
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Personalized AI-Based Interview Preparation and Evaluation System

Parikshit Jaybhaye, Hemangi Koli, Nandini Thakkar, Om Latkar, Prof. M. S. Patil
Department of Computer Engineering Pune Institute of Computer Technology
Pune, India

Abstract - In the contemporary competitive job market, ef- fective interview preparation is crucial for candidate success. However, traditional methods often lack personalization, realism, and comprehensive feedback. This survey provides a compre- hensive overview of AI-powered interview preparation systems that leverage Natural Language Processing (NLP) and Machine Learning (ML) to offer personalized mock interviews and de- tailed performance analytics. We examine the key components of these systems, including resume parsing, adaptive question gener- ation, multimodal response evaluation (encompassing knowledge, behavioral cues, and speech analysis), and the provision of action- able feedback. The study explores various technical approaches, such as semantic matching for question selection, Large Language Models (LLMs) for dialogue generation, and advanced evaluation metrics. Furthermore, we identify open challenges and future research directions, such as enhancing multimodal analysis, improving the adaptability of AI interviewers, and addressing ethical considerations. By synthesizing recent advancements, this survey aims to elucidate the potential of AI to revolutionize interview training, making it more accessible, effective, and aligned with real-world requirements.

Index Terms - Interview Preparation, Artificial Intelligence, Natural Language Processing, Question Generation, Performance Evaluation, Candidate Profiling, Machine Learning, Semantic Matching, Adaptive Learning.

I. INTRODUCTION

The job interview process is a critical gateway to career opportunities, yet it remains a significant source of anxiety and challenge for many candidates. Traditional preparation methods, such as self-study and peer-based mock interviews, often fall short in providing realistic, personalized, and objec- tive feedback. The advent of Artificial Intelligence (AI) has paved the way for innovative solutions to these challenges. AI-based interview preparation systems aim to simulate real interview environments, offer adaptive questioning based on a candidate's profile, and deliver granular feedback on both technical knowledge and soft skills.

The growing reliance on AI for professional training and assessment has resulted in a proliferation of research focused on intelligent tutoring systems. Studies indicate that the ef- fectiveness of interview preparation is significantly enhanced by personalized and interactive practice [1], [2]. For instance, previous work on mock interview systems has demonstrated the efficacy of using NLP techniques to parse resumes and generate relevant questions [1]. However, despite these ad- vancements, a notable research gap exists in the holistic in- tegration of candidate profiling, adaptive dialogue generation, and multimodal performance evaluation into a single, seamless platform.

To address this gap, this survey offers a comprehensive examination of the technologies underpinning modern AI- driven interview systems. The integration of resume parsing and semantic matching, as suggested by research on Retrieval-Augmented Generation (RAG) frameworks, can enable highly personalized question selection [4], [5]. This targeted approach ensures that the interview practice is directly relevant to the candidate's experience and the target job role.

Moreover, the deployment of AI interviewers powered by Large Language Models (LLMs) can provide a dynamic and realistic interaction experience [2], [3]. Existing literature highlights the potential of LLMs to generate contextually appropriate questions and evaluate responses in a nuanced manner [2]. By utilizing insights from these studies, our survey analyzes the architectural choices that lead to effective and engaging interview simulations.

The incorporation of multimodal evaluation—analyzing speech, text, and visual cues—further enriches these systems by providing candidates with comprehensive feedback on their performance. Research on automated interview analysis has shown that features such as prosody, facial expressions, and language fluency are strong predictors of interview outcomes [9], [10]. By synthesizing these findings, this survey not only outlines the current state-of-the-art but also contributes to a clearer roadmap for future development in the field.

This research is motivated by the need to make high-quality interview preparation more accessible and effective. By exploring the convergence of AI, NLP, and ML techniques, we aim to provide a structured overview of how intelligent systems can transform candidate readiness for the modern job market.

II. RELATED WORK

The domain of AI-assisted interview preparation has evolved rapidly, drawing from advancements in natural lan- guage processing, dialogue systems, and affective computing. While early systems focused primarily on question-answer banks, contemporary research aims to create end-to-end, in- teractive, and adaptive experiences. This section reviews the key technological strands that constitute modern AI interview systems, categorizing them into core components: Question Generation and Personalization, AI Interviewer and Dialogue Management, and Multimodal Performance Evaluation.

Question Generation and Personalization

The foundation of a personalized mock interview lies in generating relevant questions tailored to the candidate's re- sume and target job description. Early approaches relied on static databases, which lacked adaptability. Recent systems leverage web scraping and semantic matching to dynamically source and select questions.

Semantic Matching and Retrieval-Augmented Generation (RAG): Techniques like cosine similarity and transformer- based embeddings (e.g., BERT, SBERT) are used to match candidate resume entities with a knowledge base of interview questions [1]. The Retrieval-Augmented Generation (RAG) paradigm enhances this by allowing systems to retrieve rel- evant information from external corpora before generating questions. However, standard RAG can retrieve irrelevant doc- uments. Frameworks like REAR [5] address this by integrating an explicit relevance assessment module within the LLM, leading to more precise retrieval. Similarly, Self-RAG [4] introduces a critique mechanism where the model reflects on the utility of retrieved documents, improving the quality and factuality of generated content.

Resume Parsing and Candidate Profiling: Accurate ex- traction of skills, experience, and projects from resumes is crucial. NLP techniques, including Named Entity Recognition (NER) and relation extraction, are employed to build a struc- tured candidate profile. This profile is then used to steer the question generation process towards the candidate's specific background, ensuring the interview's relevance [1], [2].

AI Interviewer and Dialogue Management

The core interactive component is the AI interviewer, which must conduct a fluid, context-aware conversation. This capa- bility is largely driven by advances in Large Language Models. Large Language Models (LLMs) for Simulation: LLMs like GPT-4 and PaLM-2 form the backbone of modern dialogue systems. The MockLLM framework [2] exemplifies this by using a single LLM to simulate multiple roles (interviewer and candidate) through a multi-role, multi-behavior design. It em- ploys reflection memory and dynamic prompt modification to maintain context throughout the interview, generating realistic follow-up questions. Conversate [3] further enhances this by supporting a reflective feedback loop, allowing candidates to discuss and refine their answers with the AI after the interview. Adaptive Learning: The interview should adapt its difficulty and focus based on the candidate's performance. While fully adaptive pipelines are still an area of active research, initial systems modify subsequent questions based on the correctness and depth of previous answers, moving from general to specific or from simple to complex topics [3].

Multimodal Performance Evaluation

A key differentiator of advanced systems is their ability to evaluate performance beyond textual content, analyzing verbal and non-verbal cues. Speech and Prosody Analysis: Automatic Speech Recog- nition (ASR) is a critical first step for evaluating spoken responses. Models like Whisper [6] and AudioPaLM [7] pro- vide robust, multilingual transcription, which is then analyzed for fluency, filler words, and clarity. Prosodic features such as pitch, pace, and pauses are extracted and correlated with confidence levels [9], [10].

Visual and Behavioral Analysis: Computer Vision tech- niques are used to analyze video feeds of candidates. Convolu- tional Neural Networks (CNNs) can detect facial expressions (e.g., smiling, nervousness) and body language, which are in- dicators of confidence and engagement [10]. The study by [9] successfully predicted traits like friendliness and excitement using such features.

Knowledge and Content Evaluation: The semantic content of transcribed answers is evaluated for technical accuracy, depth, and relevance. This can involve keyword matching, se- mantic similarity against ideal answers, or more sophisticated LLM-based critique mechanisms that assess logical coherence and completeness [3], [4].

While previous studies have made significant strides in individual components, our survey synthesizes these efforts to present a holistic view of the integrated pipeline required for an effective AI-based interview preparation system.

III. PROPOSED APPROACH

This section outlines a comprehensive framework for a Personalized AI-Based Interview Preparation and Evaluation System. The proposed architecture integrates multiple AI com- ponents to create an end-to-end solution that begins with re- sume parsing and concludes with detailed feedback generation. The system workflow is designed to be adaptive, multimodal, and highly personalized.

System Architecture Overview

The proposed system follows a modular architecture con- sisting of four main components: Resume Parser and Candidate Profiler, (2) Adaptive Question Generator, (3) Mul- timodal Interview Conductor, and (4) Performance Analyzer and Feedback Generator. These components work in tandem to simulate a realistic interview experience.

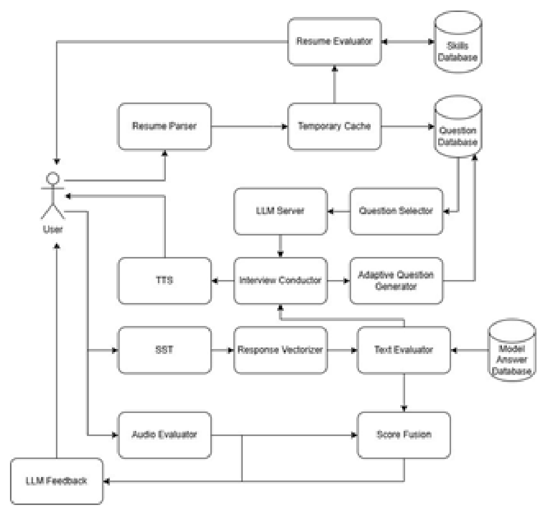

Fig. 1. Proposed system architecture for AI-based interview preparation.

Resume Parsing and Candidate Profiling

The system initiates by processing the candidate's uploaded resume. We employ a transformer-based Named Entity Recog- nition (NER) model fine-tuned on resume data to extract key entities such as Skills, Projects, Work Experience, and Education. The extracted information is structured into a candidate profile vector Pc.

The similarity between the candidate's profile and a question q in the database is computed using cosine similarity on Sentence-BERT embeddings:

$$\mathrm{sim}(P_c, q) = \frac{\phi(P_c) \cdot \phi(q)}{\|\phi(P_c)\| \, \|\phi(q)\|}$$

where ϕ is the embedding function. Questions exceeding a threshold τ = 0.7 are selected for the personalized question bank.

Adaptive Question Generation

The question generation module employs a two-stage pro- cess: retrieval and generation. First, a retrieval mechanism fetches relevant question templates based on the candidate's profile. Subsequently, a fine-tuned Large Language Model (e.g., GPT-3.5 Turbo) generates contextually varied questions. The generation is conditioned on the conversation history Ht to ensure coherence and avoid repetition.

The probability distribution for generating the next question
qt+1 is given by:

$$P(q_{t+1} | P_c, H_t) = \prod_{i=1} P(w_i | w_{<i}, P_c, H_t)$$

The system adapts question difficulty based on real-time performance assessment, escalating complexity for correct answers and providing simpler, guiding questions for incorrect responses.

Dual-Modality Response Evaluation

Candidate responses are evaluated across two modalities: audio and its transcribed text. This dual-pronged approach allows for a holistic assessment of both what the candidate said and how they said it.

Speech and Confidence Analysis: The audio stream is first processed by OpenAI's Whisper model [6] for tran- scription. The raw audio is then analyzed to extract prosodic features that are strong predictors of interview performance [9], such as speaking rate, pause frequency, and vocal pitch variance. A dedicated classifier, inspired by the work of Man- dal et al. [10], also analyzes these audio signals to determine the candidate's real-time confidence level.

Content and Accuracy Analysis: The transcribed text undergoes a two-part evaluation:
Lexical and Fluency Analysis: The transcript is ana-lyzed for key lexical features proven to correlate with positive outcomes, including the use of collaborative language ("we" vs. "I"), the number of unique words, and the frequency of filler words (e.g., "umm") [9].
Factual Accuracy Verification: To assess the substance of the answer, the system compares the candidate's claims against their pre-parsed resume. Using a critique mechanism inspired by SELF-RAG [4], the system veri- fies if the statements are [fully supported], [partially sup- ported], or [contradictory] to the information provided in the resume.

Feedback Generation

The final feedback is a comprehensive report integrating all evaluation metrics. The system moves beyond generic advice by generating specific, evidence-backed recommendations de- rived from its analysis and foundational research [9]. It uses an LLM to translate quantitative scores into actionable, natural language advice.
Example 1 (Fluency): If a high number of filler words is detected, the feedback could be: "Our analysis noted several filler words. Research indicates that speaking more fluently is perceived positively; try pausing silently to gather your thoughts" [9].

Example 2 (Content): If an answer lacks detail from the resume, the feedback might be: "Your description of Project X was good. To make it stronger, consider including the specific technologies you listed on your resume for that project, as this demonstrates deeper expertise."

Key Technologies and Methodologies

This section delves into the core technical components that enable the functionality of AI-driven interview systems. We discuss the relevant mathematics, algorithms, and models

that underpin resume parsing, question generation, dialogue management, and multimodal evaluation.

Resume Parsing and Semantic Matching

The process begins with extracting structured information from an unstructured resume. This involves Named Entity Recognition (NER) to identify entities like Skills, Organiza- tions, Degrees, and Projects.

Once a candidate profile Pc is built, it is matched against a database of job descriptions J and interview questions Q. Semantic similarity is computed using embedding models. Let ϕ(x) be a function that maps a text snippet x to a dense vector embedding (e.g., using Sentence-BERT). The similarity between a candidate's skill sc and a question q can be computed using cosine similarity:

$$\mathrm{sim}(s_c, q) = \frac{\phi(s_c) \cdot \phi(q)}{\|\phi(s_c)\| \, \|\phi(q)\|}$$

Questions with a similarity score above a threshold τ are selected for the interview. This ensures personalization based on the candidate's unique profile.

Adaptive Question Generation using LLMs

Large Language Models are fine-tuned to generate interview questions. The generation process can be conditioned on the candidate profile Pc and the conversation history Ht to ensure adaptability.

The probability of generating a question qt+1 given the context is modeled as:

$$P(q_{t+1}|P_c, H_t) = \prod_{i=1}^{L} P(w_i|w_{<i}, P_c, H_t)$$

where L is the length of the question and wi are the tokens. Models like GPT-4 and PaLM-2 are adept at this conditional generation, creating coherent and contextually relevant ques- tions [2].

Dual-Modality Response Evaluation

Candidate evaluation is a multi-faceted process involving the fusion of scores from audio and text analysis.

Speech Analysis: After transcription with an ASR model like Whisper [6], prosodic features are extracted. Met- rics such as speaking rate (R), pause frequency (Fp), and the standard deviation of fundamental frequency (σF 0) are used to build a model that classifies confidence and engagement levels [9], [10].

Content Analysis: The transcribed text (Tc) is eval- uated for quality. A language model scores the answer on criteria like accuracy and completeness by comparing it to information extracted from the candidate's resume (Pc). The model can be trained to output structured critique tags (e.g., [supported], [contradictory]) to quantify the factual alignment between the answer and the resume [4].

Overall Score Fusion: The final evaluation score (Stotal) is a weighted combination of the individual scores from the knowledge and speech analysis modules:

$$S_{\text{total}} = w_k \cdot S_{\text{knowledge}} + w_s \cdot S_{\text{speech}}$$

Each term in the formula is defined as follows:
Stotal is the final, comprehensive score for the candidate's performance on a given question. Sknowledge represents the score for the content of the answer, assessing its factual accuracy and completeness based on the transcribed text.

Sspeech is the score for the delivery of the answer, derived from analyzing audio features like prosody to evaluate confidence and fluency.

wk is the knowledge weight, a coefficient that determines the relative importance of the content score (Sknowledge) in the final calculation.

ws is the speech weight, a coefficient that determines the elative importance of the delivery score (Sspeech) in the final calculation.

where wk and ws are weights determined empirically or through machine learning. This formula is adapted from broader multimodal frameworks [9] but specialized for a voice-only context by excluding visual components.

Evaluation Metrics for System Performance

The performance of the AI system itself is measured using standard NLP and ML metrics.

- Question Relevance: Measured using Precision, Recall, and F1-score against a gold standard of relevant questions for a given profile.

- Answer Quality Evaluation: The correlation between AI-generated scores and human expert scores can be mea- sured using metrics like Pearson correlation coefficient.
- ASR Performance: Word Error Rate (WER) is the standard metric for speech recognition accuracy.
- Dialogue Quality: Metrics like BLEU or ROUGE might be used, but user satisfaction surveys are often more meaningful for interactive systems.

Challenges and Future Directions

Despite significant progress, several challenges remain in developing robust and widely applicable AI interview systems.

Handling Multimodality and Context

A major challenge is the seamless integration of text, audio, and visual cues to form a coherent understanding of candi- date performance. Current systems often analyze modalities in isolation. Future work should focus on developing fused multimodal models that can understand, for instance, how speech disfluencies correlate with facial expressions to indicate nervousness. Architectures like AudioPaLM [7], which unify speech and text processing, point towards this direction.

Adaptability and Personalization Depth

While current systems personalize questions based on re- sumes, future systems need to adapt in real-time to the candidate's emotional state, confidence level, and knowledge depth. This requires more sophisticated reinforcement learning frameworks where the AI interviewer dynamically adjusts its strategy to optimally challenge and assess the candidate.

Ethical Considerations and Bias Mitigation

AI systems can perpetuate biases present in their training data, leading to unfair assessments based on gender, accent, or ethnicity. Ensuring fairness and transparency is paramount. Future research must develop robust debiasing techniques and create diverse, inclusive datasets for training and evaluation. Explainable AI (XAI) methods will be crucial for building trust, allowing candidates to understand the rationale behind their feedback.

Realism and Candidate Experience

Creating a truly realistic interview simulation that reduces the "uncanny valley" effect remains a challenge. Integrating more advanced Text-to-Speech (TTS) systems like CMDTTS [8] for the AI's voice and generating non-verbal behaviors for AI avatars can enhance realism. Furthermore, simulating panel interviews or incorporating complex behavioral questions will better prepare candidates for real-world scenarios.

Data Scarcity and Generalization

There is a scarcity of high-quality, publicly available in- terview dialogue datasets, especially for specialized domains. Techniques for data augmentation, transfer learning, and few- shot learning will be essential for building systems that gen- eralize across different job roles and industries.

IV. CONCLUSION

This survey has provided a comprehensive overview of the emerging field of AI-powered interview preparation and evaluation systems. We have detailed the key technological components, including resume parsing, semantic question gen- eration, LLM-driven dialogue management, and multimodal performance analysis. By reviewing the current literature, we have highlighted the significant potential of these systems to make interview practice more accessible, personalized, and effective.

The integration of technologies from NLP, speech process- ing, and computer vision enables a holistic evaluation that sur- passes traditional methods. However, as outlined, challenges related to multimodal fusion, deep personalization, ethical bias, and realism remain active areas of research. Addressing these challenges will require collaborative efforts from the AI community.

The future of interview preparation is poised to be increas- ingly driven by AI. As these systems become more sophisti- cated and widespread, they have the potential to democratize access to high-quality career coaching, ultimately leading to a better-prepared workforce and a more efficient hiring process. Continued research and development in this area are essential to fully realize this potential.

REFERENCES

1. EZInterviewer: To Improve Job Interview Performance with Mock Interview Generator.
2. Facilitating Multi-Role and Multi-Behavior Collaboration of Large Lan- guage Models for Online Job Seeking and Recruiting.
3. Conversate: Supporting Reflective Learning in Interview Practice Through Interactive Simulation and Dialogic Feedback.
4. Asai et al., "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection," 2023.
5. REAR: A Relevance-Aware Retrieval-Augmented Framework.
6. Radford et al., "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI, 2023.
7. Kharitonov et al., "AudioPaLM: A Large Language Model That Can Speak and Listen," Google, 2023.
8. CMDTTS: A Fast and High-quality TTS Method with Compressed Corpus.
9. Automated Analysis and Prediction of Job Interview Performance.
10. R. Mandal et al., "AI -Based mock interview evaluator: An emotion and confidence classifier model," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS).